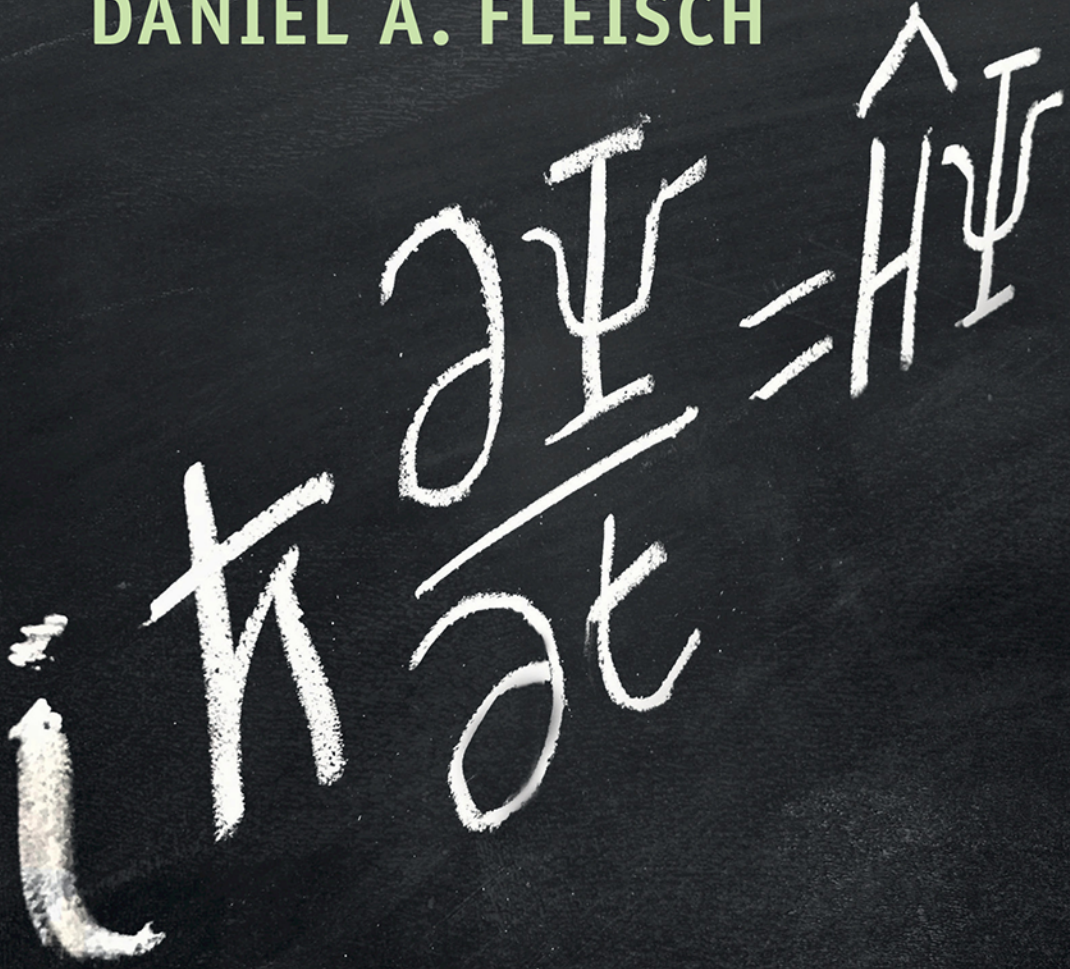
A Student's Guide to the Schrödinger Equation

DANIEL A. FLEISCH



CAMBRIDGE

A Student's Guide to the Schrödinger Equation

Quantum mechanics is a hugely important topic in science and engineering, but many students struggle to understand the abstract mathematical techniques used to solve the Schrödinger equation and to analyze the resulting wave functions. Retaining the popular approach used in Fleisch's other Student's Guides, this friendly resource uses plain language to provide detailed explanations of the fundamental concepts and mathematical techniques underlying the Schrödinger equation in quantum mechanics. It addresses in a clear and intuitive way the problems students find most troublesome. Each chapter includes several homework problems with fully worked solutions. A companion website hosts additional resources, including a helpful glossary, Matlab code for creating key simulations, revision quizzes and a series of videos in which the author explains the most important concepts from each section of the book.

DANIEL A. FLEISCH is Emeritus Professor of Physics at Wittenberg University, where he specializes in electromagnetics and space physics. He is the author of four other books published by Cambridge University Press: *A Student's Guide to Maxwell's Equations* (2008); *A Student's Guide to Vectors and Tensors* (2011); *A Student's Guide to the Mathematics of Astronomy* (2013); and *A Student's Guide to Waves* (2015).

Other books in the Student's Guide series:

A Student's Guide to General Relativity, Norman Gray

A Student's Guide to Analytical Mechanics, John L. Bohn

A Student's Guide to Infinite Series and Sequences, Bernhard W. Bach Jr.

A Student's Guide to Atomic Physics, Mark Fox

A Student's Guide to Waves, Daniel A. Fleisch, Laura Kinnaman

A Student's Guide to Entropy, Don S. Lemons

A Student's Guide to Dimensional Analysis, Don S. Lemons

A Student's Guide to Numerical Methods, Ian H. Hutchinson

A Student's Guide to Langrangians and Hamiltonians, Patrick Hamill

A Student's Guide to the Mathematics of Astronomy, Daniel A. Fleisch, Julia Kregenow

A Student's Guide to Vectors and Tensors, Daniel A. Fleisch

A Student's Guide to Maxwell's Equations, Daniel A. Fleisch

A Student's Guide to Fourier Transforms, J. F. James

A Student's Guide to Data and Error Analysis, Herman J. C. Berendsen

A Student's Guide to the Schrödinger Equation

DANIEL A. FLEISCH
Wittenberg University

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre,
New Delhi – 110025, India

79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781108834735
DOI: 10.1017/9781108834735

© Daniel A. Fleisch 2020

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2020

Printed and bound in Great Britain by Clays Ltd, Elcograf S.p.A.

A catalogue record for this publication is available from the British Library.

Library of Congress Cataloging-in-Publication Data
Names: Fleisch, Daniel A., author.
Title: A student's guide to the Schrödinger equation / Daniel A. Fleisch (Wittenberg University, Ohio).
Other titles: Schrödinger equation
Description: Cambridge ; New York, NY : Cambridge University Press, 2020. | Includes bibliographical references and index.
Identifiers: LCCN 2018035530 | ISBN 9781108834735 (hardback) | ISBN 9781108819787 (pbk.)
Subjects: LCSH: Schrödinger equation–Textbooks. | Quantum theory.
Classification: LCC QC174.26.W28 F5545 2020 | DDC 530.12/4–dc23
LC record available at https://lccn.loc.gov/2018035530

ISBN 978-1-108-83473-5 Hardback
ISBN 978-1-108-81978-7 Paperback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

About this book

This edition of *A Student's Guide to the Schrödinger Equation* is supported by an extensive range of interactive digital resources, available via a companion website. These resources have been designed to support your learning and bring the textbook to life, supporting active learning and providing you with feedback.

Please visit **www.cambridge.org/fleisch-SGSE** to access this extra content.

The following icons appear throughout the book in the bottom margin and indicate where resources for that page are available on the website.

Interactive Simulation

Learning Objective

Video

Worked Problem

Quiz

Glossary - Glossary items are highlighted bold in the text and full explanations of the term can be found on the website

We may update our Site from time to time, and may change or remove the content at any time. We do not guarantee that our Site, or any part of it, will always be available or be uninterrupted or error free. Access to our Site is permitted on a temporary and "as is" basis. We may suspend or change all or any part of our Site without notice. We will not be liable to you if for any reason our Site or the content is unavailable at any time, or for any period.

Contents

	Preface	*page* ix
	Acknowledgments	xi
1	**Vectors and Functions**	1
1.1	Vector Basics	2
1.2	Dirac Notation	8
1.3	Abstract Vectors and Functions	14
1.4	Complex Numbers, Vectors, and Functions	18
1.5	Orthogonal Functions	22
1.6	Finding Components Using the Inner Product	26
1.7	Problems	30
2	**Operators and Eigenfunctions**	32
2.1	Operators, Eigenvectors, and Eigenfunctions	32
2.2	Operators in Dirac Notation	37
2.3	Hermitian Operators	43
2.4	Projection Operators	49
2.5	Expectation Values	56
2.6	Problems	60
3	**The Schrödinger Equation**	63
3.1	Origin of the Schrödinger Equation	64
3.2	What the Schrödinger Equation Means	71
3.3	Time-Independent Schrödinger Equation	78
3.4	Three-Dimensional Schrödinger Equation	81
3.5	Problems	93

4 Solving the Schrödinger Equation 95
4.1 The Born Rule and Copenhagen Interpretation 96
4.2 Quantum States, Wavefunctions, and Operators 98
4.3 Characteristics of Quantum Wavefunctions 102
4.4 Fourier Theory and Quantum Wave Packets 111
4.5 Position and Momentum Wavefunctions and Operators 132
4.6 Problems 144

5 Solutions for Specific Potentials 146
5.1 Infinite Rectangular Potential Well 147
5.2 Finite Rectangular Potential Well 168
5.3 Harmonic Oscillator 193
5.4 Problems 216

References 218
Index 219

Preface

This book has one purpose: to help you understand the Schrödinger equation and its solutions. Like my other *Student's Guides*, this book contains explanations written in plain language and supported by a variety of freely available online materials. Those materials include complete solutions to every problem in the text, in-depth discussions of supplemental topics, and a series of video podcasts in which I explain the most important concepts, equations, graphs, and mathematical techniques of every chapter.

This *Student's Guide* is intended to serve as a supplement to the many comprehensive texts dealing with the Schrödinger equation and quantum mechanics. That means that it's designed to provide the conceptual and mathematical foundation on which your understanding of quantum mechanics will be built. So if you're enrolled in a course in quantum mechanics, or you're studying modern physics on your own, and you're not clear on the relationship between wave functions and vectors, or you want to know the physical meaning of the inner product, or you're wondering exactly what eigenfunctions are and why they're so important, then this may be the book for you.

I've made this book as modular as possible to allow you to get right to the material in which you're interested. Chapters 1 and 2 provide an overview of the mathematical foundation on which the Schrödinger equation and the science of quantum mechanics is built. That includes generalized vector spaces, orthogonal functions, operators, eigenfunctions, and the Dirac notation of bras, kets, and inner products. That's quite a load of mathematics to work through, so in each section of those two chapters you'll find a "Main Ideas" statement that concisely summarizes the most important concepts and techniques of that section, as well as a "Relevance to Quantum Mechanics"

paragraph that explains how that bit of mathematics relates to the physics of quantum mechanics.

So I recommend that you take a look at the "Main Ideas" statements in each section of Chapters 1 and 2, and if your understanding of those topics is solid, you can skip past that material and move right into a term-by-term dissection of the Schrödinger equation in both time-dependent and time-independent form in Chapter 3. And if you're confident in your understanding of the meaning of the Schrödinger equation, you can dive into Chapter 4, in which you'll find a discussion of the quantum wavefunctions that are solutions to that equation. Finally, in Chapter 5, you can see how these principals and mathematical techniques are applied to three situations with specific potentials: the infinite rectangular potential well, the finite rectangular potential well, and the quantum harmonic oscillator.

As I hope you can tell, I spend a lot of time thinking about the best way to explain challenging concepts that my students find troubling. My *Student's Guides* are the result of that thinking, and my goal in writing them is elegantly expressed by A. W. Sparrow in his wonderful little book *Basic Wireless*: "This booklet makes no pretence of superseding the numerous textbooks already published. It hopes to prove a convenient stepping-stone towards them by concise presentation of foundation knowledge." If my efforts are half as successful as those of Sparrow, you should find this book helpful.

Acknowledgments

If you find the explanations in this *Student's Guide* helpful, it's because of the insightful questions and helpful feedback I've received from the students in my Physics 411 (Quantum Mechanics) course at Wittenberg University. Their willingness to take on the formidable challenge of understanding abstract vector spaces, eigenvalue equations, and quantum operators has provided the inspiration to keep me going when the going got, let's say, "uncertain." I owe them a lot.

Thanks is also due to Dr. Nick Gibbons, Dr. Simon Capelin, and the production team at Cambridge University Press for their professionalism and steady support during the planning, writing, and production of this book.

Most curiously, after five *Student's Guides*, twenty years of teaching, and an increasing fraction of our house taken over by physics books, astronomical instrumentation, and draft manuscripts, Jill Gianola continues to encourage my efforts. For that, I have no explanation.

1

Vectors and Functions

There's a great deal of interesting physics in the Schrödinger equation and its solutions, and the mathematical underpinnings of that equation can be expressed in several ways. It's been my experience that students find it helpful to see a combination of Erwin Schrödinger's wave mechanics approach and the matrix mechanics approach of Werner Heisenberg, as well as Paul Dirac's bra and ket notation. So these first two chapters provide the mathematical foundations that will help you understand these different perspectives and "languages" of quantum mechanics, beginning with the basics of vectors in Section 1.1. With that basis in place, you can move on to Dirac notation in Section 1.2 and abstract vectors and functions in Section 1.3. The rules pertaining to complex numbers, vectors, and functions are reviewed in Section 1.4, followed by an explanation of orthogonal functions in Section 1.5, and using the inner product to find components in Section 1.6. The final section of this chapter (as in all later chapters) is a set of problems that will allow you to exercise your understanding of the concepts and mathematical techniques presented in this chapter. Remember that you can find full, interactive solutions to every problem on the book's website.

And since it's easy to lose sight of the architectural plan of an elaborate structure when you're laying the foundation, as mentioned in the Preface you'll find in each section a plain-language statement of the main ideas of that section as well as a short paragraph explaining the relevance of that development to the Schrödinger equation and quantum mechanics.

As you look through this chapter, don't forget that this book is modular, so if you have a good understanding of the included topics and their relevance to quantum mechanics, you should feel free to skip over this chapter and jump into the discussions of operators and eigenfunctions in Chapter 2. And if you're

already up to speed on those topics, the Schrödinger equation and quantum wavefunctions await your attention in later chapters.

1.1 Vector Basics

If you pick up any book about quantum mechanics, you're sure to find lots of discussion about wavefunctions and the solutions to the Schrödinger equation. But the language used to describe those functions, and the mathematical techniques used to analyze them, are rooted in the world of **vectors**. I've noticed that students who have a thorough understanding of basis vectors, inner products, and vector components are far more likely to succeed when they encounter the more advanced aspects of quantum mechanics, so this section is all about vectors.

When you first learned about vectors, you probably thought of a vector as an entity that has both magnitude (length) and direction (angles from some set of axes). You may also have learned to write a vector as a letter with a little arrow over its head (such as $\vec{A}$), and to "expand" a vector like this:

$$\vec{A} = A_x\hat{\imath} + A_y\hat{\jmath} + A_z\hat{k}. \tag{1.1}$$

In this expansion, A_x, A_y, and A_z are the components of vector $\vec{A}$, and $\hat{\imath}$, $\hat{\jmath}$, and $\hat{k}$ are directional indicators called "basis vectors" of the coordinate system you're using to expand vector $\vec{A}$. In this case, that's the Cartesian (x, y, z) coordinate system shown in Fig. 1.1. It's important to understand that vector $\vec{A}$ exists independently of any particular basis system; the same vector may be expanded in many different basis systems.

The basis vectors $\hat{\imath}$, $\hat{\jmath}$, and $\hat{k}$ are also called "unit vectors" because they each have length of one unit. And what unit is that? Whatever unit you're using to express the length of vector $\vec{A}$. It may help you to think of a unit vector as defining one "step" along a coordinate axis, so an expression such as

$$\vec{A} = 5\hat{\imath} - 2\hat{\jmath} + 3\hat{k}, \tag{1.2}$$

tells you to take five steps in the (positive) x-direction, two steps in the (negative) y-direction, and three steps in the (positive) z-direction to get from the start to the end of the vector $\vec{A}$.

You may also recall that the magnitude (that is, the length or "norm") of a vector, usually written as $|\vec{A}|$ or $\|\vec{A}\|$, can be found from its Cartesian components using the equation

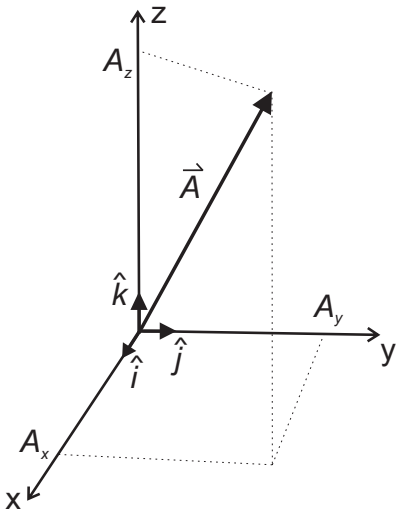


Figure 1.1 Vector $\vec{A}$ with its Cartesian components A_x, A_y, and A_z and the Cartesian unit vectors $\hat{\imath}$, $\hat{\jmath}$, and $\hat{k}$.

$$|\vec{A}| = \sqrt{A_x^2 + A_y^2 + A_z^2}, \tag{1.3}$$

and that the negative of a vector (such as $-\vec{A}$) is a vector of the same length as $\vec{A}$ but pointed in the opposite direction.

Adding two vectors together can be done graphically, as shown in Fig. 1.2, by sliding one vector (without changing its direction or length) so that its tail is at the head of the other vector; the sum is a new vector drawn from the tail of the undisplaced vector to the head of the displaced vector. Alternatively, vectors may be added analytically by adding the components in each direction:

$$\begin{aligned} \vec{A} &= A_x\hat{\imath} + A_y\hat{\jmath} + A_z\hat{k} \\ +\ \vec{B} &= B_x\hat{\imath} + B_y\hat{\jmath} + B_z\hat{k} \\ \vec{C} = \vec{A}+\vec{B} &= (A_x + B_x)\hat{\imath} + (A_y + B_y)\hat{\jmath} + (A_z + B_z)\hat{k}. \end{aligned} \tag{1.4}$$

Another important operation is multiplying a vector by a scalar (that is, a number with no directional indicator), which changes the length but not the direction of the vector. So if α is a scalar, then

$$\begin{aligned} \vec{D} = \alpha\vec{A} &= \alpha(A_x\hat{\imath} + A_y\hat{\jmath} + A_z\hat{k}) \\ &= \alpha A_x\hat{\imath} + \alpha A_y\hat{\jmath} + \alpha A_z\hat{k}. \end{aligned}$$

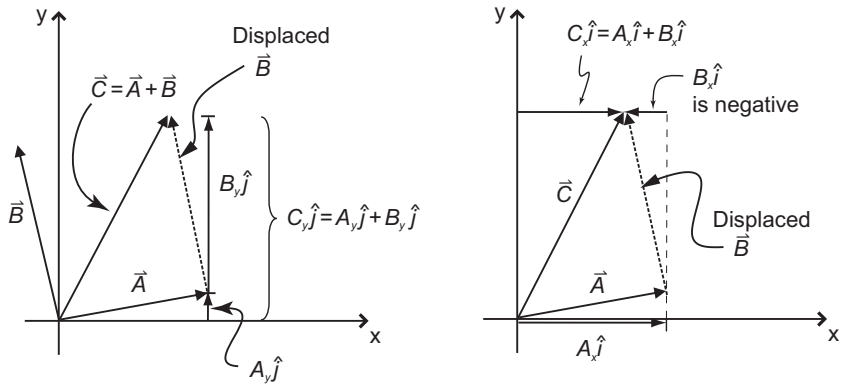


Figure 1.2 Adding vectors $\vec{A}$ and $\vec{B}$ graphically by sliding the tail of vector $\vec{B}$ to the head of vector $\vec{A}$ without changing its length or direction.

Scaling each component equally (by factor α) means that vector $\vec{D}$ points in the same direction as $\vec{A}$, but the length of $\vec{D}$ is

$$\begin{aligned}|\vec{D}| &= \sqrt{D_x^2 + D_y^2 + D_z^2} \\ &= \sqrt{(\alpha A_x)^2 + (\alpha A_y)^2 + (\alpha A_z)^2} \\ &= \sqrt{\alpha^2(A_x^2 + A_y^2 + A_z^2)} = \alpha|\vec{A}|.\end{aligned}$$

So the vector's length is scaled by the factor α, but its direction remains the same (unless α is negative, in which case the direction reverses, but the vector still lies along the same line).

Relevance to Quantum Mechanics

As you'll see in later chapters, the solutions to the Schrödinger equation are quantum wavefunctions that behave like generalized higher-dimensional vectors. That means they can be added together to form a new wavefunction and they can be multiplied by scalars without changing their "direction." How functions can have "length" and "direction" is explained in Chapter 2.

In addition to summing vectors, multiplying vectors by scalars, and finding the length of vectors, another important operation is the scalar[1] product

[1] Note that this is called the scalar product because the result is a scalar, not because a scalar is involved in the multiplication.

(also called the "dot product") of two vectors, usually written as $(\vec{A}, \vec{B})$ or $\vec{A} \circ \vec{B}$. The scalar product is given by

$$(\vec{A}, \vec{B}) = \vec{A} \circ \vec{B} = |\vec{A}||\vec{B}| \cos\theta, \tag{1.5}$$

in which θ is the angle between $\vec{A}$ and $\vec{B}$. In Cartesian coordinates, the dot product may be found by multiplying corresponding components and summing the results:

$$(\vec{A}, \vec{B}) = \vec{A} \circ \vec{B} = A_x B_x + A_y B_y + A_z B_z. \tag{1.6}$$

Notice that if vectors $\vec{A}$ and $\vec{B}$ are parallel, then the dot product is

$$\vec{A} \circ \vec{B} = |\vec{A}||\vec{B}| \cos\left(0^\circ\right) = |\vec{A}||\vec{B}|, \tag{1.7}$$

since $\cos(0^\circ) = 1$. Alternatively, if $\vec{A}$ and $\vec{B}$ are perpendicular, then the value of the dot product is zero:

$$\vec{A} \circ \vec{B} = |\vec{A}||\vec{B}| \cos\left(90^\circ\right) = 0, \tag{1.8}$$

since $\cos(90^\circ) = 0$.

The dot product of a vector with itself gives the square of the magnitude of the vector:

$$\vec{A} \circ \vec{A} = |\vec{A}||\vec{A}| \cos\left(0^\circ\right) = |\vec{A}|^2. \tag{1.9}$$

A generalized version of the scalar product called the "inner product" is extremely useful in quantum mechanics, so it's worth a bit of your time to think about what happens when you perform an operation such as $\vec{A} \circ \vec{B}$. As you can see in Fig. 1.3a, the term $|\vec{B}| \cos\theta$ is the projection of vector $\vec{B}$ onto the direction of vector $\vec{A}$, so the dot product gives an indication of "how much" of $\vec{B}$ lies along the direction of $\vec{A}$.[2] Alternatively, you can isolate the $|\vec{A}| \cos\theta$ portion of the dot product $\vec{A} \circ \vec{B} = |\vec{A}||\vec{B}| \cos\theta$, which is the projection of $\vec{A}$ onto the direction of $\vec{B}$, as shown in Fig. 1.3b. From this perspective, the dot product indicates "how much" of vector $\vec{A}$ lies along the direction of $\vec{B}$. Either way, the dot product provides a measure of how much one vector "contributes" to the direction of another.

To make this concept more specific, consider what you get by dividing the dot product by the magnitude of $\vec{A}$ times the magnitude of $\vec{B}$:

[2] If you find the phrase "lies along" troubling (since vector $\vec{A}$ and vector $\vec{B}$ lie in different directions), perhaps it will help to imagine a tiny traveler walking from the start to the end of vector $\vec{B}$, and asking "In walking along vector $\vec{B}$, how much does a traveler advance in the direction of vector $\vec{A}$?"

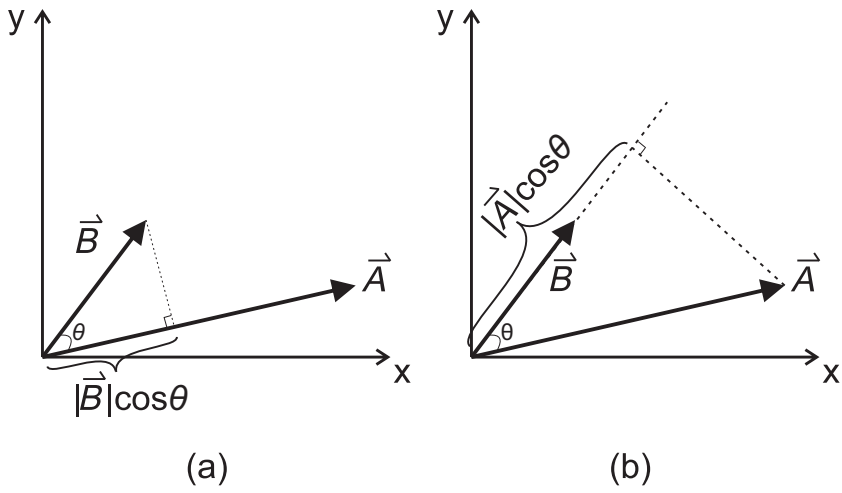


Figure 1.3 (a) The projection of vector $\vec{B}$ onto the direction of vector $\vec{A}$ and (b) the projection of vector $\vec{A}$ onto the direction of vector $\vec{B}$.

$$\frac{\vec{A} \circ \vec{B}}{|\vec{A}||\vec{B}|} = \frac{|\vec{A}||\vec{B}| \cos\theta}{|\vec{A}||\vec{B}|} = \cos\theta, \tag{1.10}$$

which ranges from one to zero as the angle between the vectors increases from 0° to 90°. So if two vectors are parallel, each contributes its entire length to the direction of the other, but if they're perpendicular, neither makes any contribution to the direction of the other.

This understanding of the dot product makes it easy to comprehend the results of taking the dot product between pairs of the Cartesian unit vectors:

Each of these unit vectors lies entirely along itself

$$\begin{cases} \hat{\imath} \circ \hat{\imath} = |\hat{\imath}||\hat{\imath}| \cos 0^\circ = (1)(1)(1) = 1 \\ \hat{\jmath} \circ \hat{\jmath} = |\hat{\jmath}||\hat{\jmath}| \cos 0^\circ = (1)(1)(1) = 1 \\ \hat{k} \circ \hat{k} = |\hat{k}||\hat{k}| \cos 0^\circ = (1)(1)(1) = 1 \end{cases}$$

No part of these unit vectors lies along any other

$$\begin{cases} \hat{\imath} \circ \hat{\jmath} = |\hat{\imath}||\hat{\jmath}| \cos 90^\circ = (1)(1)(0) = 0 \\ \hat{\imath} \circ \hat{k} = |\hat{\imath}||\hat{k}| \cos 90^\circ = (1)(1)(0) = 0 \\ \hat{\jmath} \circ \hat{k} = |\hat{\jmath}||\hat{k}| \cos 90^\circ = (1)(1)(0) = 0 \end{cases}$$

The Cartesian unit vectors are called "orthonormal" because they're orthogonal (each is perpendicular to the others) as well as normalized (each has

magnitude of one). They're also called a "**complete set**" because any vector in three-dimensional Cartesian space can be made up of a weighted combination of these three basis vectors.

Here's a very useful trick: orthonormal basis vectors make it easy to use the dot product to determine the components of a vector. For a vector $\vec{A}$, the components A_x, A_y, and A_z can be found by dotting the basis vectors $\hat{\imath}$, $\hat{\jmath}$, and $\hat{k}$ into $\vec{A}$:

$$\begin{aligned} A_x = \hat{\imath} \circ \vec{A} &= \hat{\imath} \circ (A_x\hat{\imath} + A_y\hat{\jmath} + A_z\hat{k}) \\ &= A_x(\hat{\imath} \circ \hat{\imath}) + A_y(\hat{\imath} \circ \hat{\jmath}) + A_z(\hat{\imath} \circ \hat{k}) \\ &= A_x(1) + A_y(0) + A_z(0) = A_x. \end{aligned}$$

Likewise for A_y

$$\begin{aligned} A_y = \hat{\jmath} \circ \vec{A} &= \hat{\jmath} \circ (A_x\hat{\imath} + A_y\hat{\jmath} + A_z\hat{k}) \\ &= A_x(\hat{\jmath} \circ \hat{\imath}) + A_y(\hat{\jmath} \circ \hat{\jmath}) + A_z(\hat{\jmath} \circ \hat{k}) \\ &= A_x(0) + A_y(1) + A_z(0) = A_y. \end{aligned}$$

And for A_z

$$\begin{aligned} A_z = \hat{k} \circ \vec{A} &= \hat{k} \circ (A_x\hat{\imath} + A_y\hat{\jmath} + A_z\hat{k}) \\ &= A_x(\hat{k} \circ \hat{\imath}) + A_y(\hat{k} \circ \hat{\jmath}) + A_z(\hat{k} \circ \hat{k}) \\ &= A_x(0) + A_y(0) + A_z(1) = A_z. \end{aligned}$$

This technique of digging out the components of a vector using the dot product and basis vectors is extremely valuable in quantum mechanics.

Main Ideas of This Section
Vectors are mathematical representations of quantities that may be expanded as a series of components, each of which pertains to a directional indicator called a basis vector. A vector may be added to another vector to produce a new vector, and a vector may be multiplied by a scalar or by another vector. The dot or scalar product between two vectors produces a scalar result proportional to the projection of one of the vectors along the direction of the other. The components of a vector in an orthonormal basis system may be found by dotting each basis vector into the vector.

Relevance to Quantum Mechanics
Just as a vector can be expressed as a weighted combination of basis vectors, a quantum wavefunction can be expressed as a weighted combination of basis wavefunctions. A generalized version of the dot product called the inner product can be used to calculate how much each component wavefunction contributes to the sum, and this determines the probability of various measurement outcomes.

1.2 Dirac Notation

Before making the connection between vectors and quantum wavefunctions, it's important for you to realize that vector components such as A_x, A_y, and A_z have meaning only when tied to a set of basis vectors (A_x to $\hat{\imath}$, A_y to $\hat{\jmath}$, and A_z to $\hat{k}$). If you had chosen to represent vector $\vec{A}$ using a different set of basis vectors (for example, by rotating the x-, y-, and z-axes and using basis vectors aligned with the rotated axes), you could have written the same vector $\vec{A}$ as

$$\vec{A} = A_{x'}\hat{\imath}' + A_{y'}\hat{\jmath}' + A_{z'}\hat{k}',$$

in which the rotated axes are designated x′, y′, and z′, and the basis vectors pointing along those axes are $\hat{\imath}'$, $\hat{\jmath}'$, and $\hat{k}'$.

When you expand a vector such as $\vec{A}$ in terms of different basis vectors, the vector components of the vector may change, but the new components and the new basis vectors add up to give the same vector $\vec{A}$. You may even choose to use a non-Cartesian set of basis vectors such as the spherical basis vectors $\hat{r}$, $\hat{\theta}$, and $\hat{\phi}$; expanding vector $\vec{A}$ in this basis looks like this:

$$\vec{A} = A_r\hat{r} + A_\theta\hat{\theta} + A_\phi\hat{\phi}.$$

Once again, different components, different basis vectors, but the *combination* of components and basis vectors gives the same vector $\vec{A}$.

What's the advantage of using one set of basis vectors or another? Depending on the geometry of the situation, it may be simpler to represent or manipulate vectors in a particular basis. But once you've specified a basis, a vector may be represented simply by writing its components in that basis as an ordered set of numbers.

For example, you could choose to represent a three-dimensional vector by writing its components into a single-column matrix

$$\vec{A} = \begin{pmatrix} A_x \\ A_y \\ A_z \end{pmatrix},$$

as long as you remember that vectors may be represented in this way only when the **basis system** has been specified.

Since they're vectors, the Cartesian basis vectors ($\hat{\imath}$, $\hat{\jmath}$, and $\hat{k}$) themselves can be written as column vectors. To do so, it's necessary to ask "In what basis?" Students sometimes find this a strange question, since we're talking about representing a basis vector, so isn't the basis obvious?

The answer is that it's perfectly possible to expand any vector, including a basis vector, using whichever basis system you choose. But some choices will lead to simpler representation than others, as you can see by representing $\hat{\imath}$, $\hat{\jmath}$, and $\hat{k}$ using their own Cartesian basis system:

$$\hat{\imath} = 1\hat{\imath} + 0\hat{\jmath} + 0\hat{k} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \qquad \hat{\jmath} = 0\hat{\imath} + 1\hat{\jmath} + 0\hat{k} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

and

$$\hat{k} = 0\hat{\imath} + 0\hat{\jmath} + 1\hat{k} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Such a basis system, in which each basis vector has only one nonzero component, and the value of that component is +1, is called the "standard" or "natural" basis.

Here's what it looks like if you express the Cartesian basis vectors ($\hat{\imath}, \hat{\jmath}, \hat{k}$) using the basis vectors ($\hat{r}$, $\hat{\theta}$, $\hat{\phi}$) of the spherical coordinate system

$$\begin{aligned} \hat{\imath} &= \sin\theta\cos\phi\hat{r} + \cos\theta\cos\phi\hat{\theta} - \sin\phi\hat{\phi} \\ \hat{\jmath} &= \sin\theta\sin\phi\hat{r} + \cos\theta\sin\phi\hat{\theta} + \cos\phi\hat{\phi} \\ \hat{k} &= \cos\theta\hat{r} - \sin\theta\hat{\theta}. \end{aligned}$$

So the column-vector representation of $\hat{\imath}, \hat{\jmath}, \hat{k}$ in the spherical basis system is

$$\hat{\imath} = \begin{pmatrix} \sin\theta\cos\phi \\ \cos\theta\cos\phi \\ -\sin\phi \end{pmatrix} \qquad \hat{\jmath} = \begin{pmatrix} \sin\theta\sin\phi \\ \cos\theta\sin\phi \\ \cos\phi \end{pmatrix} \qquad \hat{k} = \begin{pmatrix} \cos\theta \\ -\sin\theta \\ 0 \end{pmatrix}.$$

The bottom line is this: whenever you see a vector represented as a column of components, it's essential that you understand the basis system to which those components pertain.

Relevance to Quantum Mechanics
Like vectors, quantum wavefunctions can be expressed as a series of components, but those components have meaning only when you've defined the basis functions to which they pertain.

In quantum mechanics, you're likely to encounter entities called "ket vectors" or simply "kets," written with a vertical bar on the left and angled bracket on the right, such as $|A\rangle$. The ket $|A\rangle$ can be expanded in the same way as vector $\vec{A}$:

$$|A\rangle = A_x\,|i\rangle + A_y\,|j\rangle + A_z\,|k\rangle = \begin{pmatrix} A_x \\ A_y \\ A_z \end{pmatrix} = A_x\hat{i} + A_y\hat{j} + A_z\hat{k} = \vec{A}. \qquad (1.11)$$

So if kets are just a different way of representing vectors, why call them "kets" and write them as column vectors? This notation was developed by the British physicist Paul Dirac in 1939, while he was working with a generalized version of the dot product called the **inner product**, written as $\langle A|B\rangle$. In this context, "generalized" means "not restricted to real vectors in three-dimensional physical space," so the inner product can be used with higher-dimensional abstract vectors with complex components, as you'll see in Sections 1.3 and 1.4. Dirac realized that the inner product bracket $\langle A|B\rangle$ could be conceptually divided into two pieces, a left half (which he called a "**bra**") and a right half (which he called a "**ket**"). In conventional notation, an inner product between vectors $\vec{A}$ and $\vec{B}$ might be written as $\vec{A} \circ \vec{B}$ or $(\vec{A}, \vec{B})$, but in Dirac notation the inner product is written as

$$\text{Inner product of } |A\rangle \text{ and } |B\rangle = \langle A| \text{ times } |B\rangle = \langle A|B\rangle\,. \qquad (1.12)$$

Notice that in forming the bracket $\langle A|B\rangle$ as the multiplication of bra $\langle A|$ by ket $|B\rangle$, the right vertical bar of $\langle A|$ and the left vertical bar of $|B\rangle$ are combined into a single vertical bar.

To calculate the inner product $\langle A|B\rangle$, begin by representing vector $\vec{A}$ as a ket:

$$|A\rangle = \begin{pmatrix} A_x \\ A_y \\ A_z \end{pmatrix} \qquad (1.13)$$

in which the subscripts indicate that these components pertain to the Cartesian basis system. Now form the bra $\langle A|$ by taking the complex conjugate[3] of each component and writing them as a row vector:

$$\langle A| = \left(A_{\mathrm{x}}^* \, A_{\mathrm{y}}^* \, A_{\mathrm{z}}^*\right). \tag{1.14}$$

The inner product $\langle A|B\rangle$ is thus

$$\langle A| \text{ times } |B\rangle = \langle A|B\rangle = (A_{\mathrm{x}}^* \, A_{\mathrm{y}}^* \, A_{\mathrm{z}}^*) \begin{pmatrix} B_{\mathrm{x}} \\ B_{\mathrm{y}} \\ B_{\mathrm{z}} \end{pmatrix}. \tag{1.15}$$

By the rules of matrix multiplication, this gives

$$\langle A|B\rangle = (A_{\mathrm{x}}^* \, A_{\mathrm{y}}^* \, A_{\mathrm{z}}^*) \begin{pmatrix} B_{\mathrm{x}} \\ B_{\mathrm{y}} \\ B_{\mathrm{z}} \end{pmatrix} = A_{\mathrm{x}}^* B_{\mathrm{x}} + A_{\mathrm{y}}^* B_{\mathrm{y}} + A_{\mathrm{z}}^* B_{\mathrm{z}}, \tag{1.16}$$

as you'd expect for a generalized version of the dot product.

So kets can be represented by column vectors, and bras can be represented by row vectors, but a common question among students new to quantum mechanics is "What exactly *are* kets, and what are bras?" The answer to the first question is that kets are mathematical objects that are members of a "vector space" (also called a "linear space"). If you've studied any linear algebra, you've already encountered the concept of a **vector space**, and you may remember that a vector space is just a collection of vectors that behave according to certain rules. Those rules include the addition of vectors to produce new vectors (which live in the same space), and multiplying a vector by a scalar, producing a scaled version of the vector (which also lives in that space).

Since we'll be dealing with generalized vectors rather than vectors in three-dimensional physical space, instead of labeling the components x, y, z, we'll number them. And instead of using the Cartesian unit vectors $\hat{\imath}, \hat{\jmath}, \hat{k}$, we'll use the basis vectors $\vec{\epsilon}_1, \vec{\epsilon}_2 \ldots \vec{\epsilon}_N$. So the equation

$$|A\rangle = A_{\mathrm{x}} \, |i\rangle + A_{\mathrm{y}} \, |j\rangle + A_{\mathrm{z}} \, |k\rangle \tag{1.17}$$

becomes

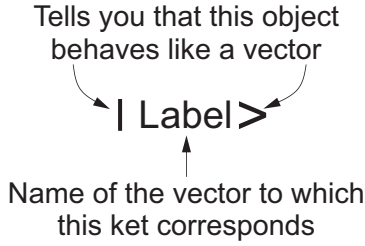
$$|A\rangle = A_1 \, |\epsilon_1\rangle + A_2 \, |\epsilon_2\rangle + \cdots A_N \, |\epsilon_N\rangle = \sum_{i=1}^{N} A_i \, |\epsilon_i\rangle \, , \tag{1.18}$$

in which A_i represents the ket component for the basis ket $|\epsilon_i\rangle$.

[3] The reason for taking the complex conjugate is explained in Section 1.4, where you'll also find a refresher on complex quantities.

But just as the vector $\vec{A}$ is the same vector no matter which coordinate system you use to express its components, the ket $|A\rangle$ exists independently of any particular set of basis kets (kets are said to be "basis independent"). So ket $|A\rangle$ behaves just like vector $\vec{A}$.

It may help you to think of a ket like this:



Once you've picked a basis system, why write the components of a ket as a column vector? One good reason is that it allows the rules of matrix multiplication to be applied to form scalar products, as in Eq. 1.16.

The other members of those scalar products are bras, and the definition of a bra is somewhat different from that of a ket. That's because a bra is a "linear functional" (also called a "**covector**" or a "one-form") that combines with a ket to produce a scalar; mathematicians say bras map vectors to the field of scalars.

So what's a linear functional? It's essentially a mathematical device (some authors refer to it as an instruction) that operates on another object. Hence a bra operates on a ket, and the result of that operation is a scalar. How does this operation map to a scalar? By following the rules of the scalar product, which you've already seen for the dot product between two real vectors. In Section 1.4 you'll learn the rules for taking the inner product between two complex abstract vectors.

Bras don't inhabit the same vector space as kets – they live in their own vector space that's called the "**dual space**" to the space of kets. Within that space, bras can be added together and multiplied by scalars to produce new bras, just as kets can in their space.

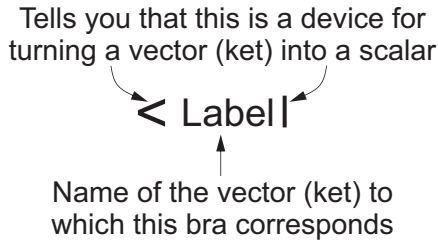
One reason that the space of bras is called "dual" to the space of kets is that for every ket there exists a corresponding bra, and when a bra operates on its corresponding (dual) ket, the scalar result is the square of the norm of the ket:

$$\langle A|A\rangle = \begin{pmatrix} A_1^* & A_2^* & \dots & A_N^* \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_N \end{pmatrix} = |\vec{A}|^2,$$

just as the dot product of a (real) vector with itself gives the square of the vector's length (Eq. 1.9).

Note that the bra that is the dual of ket $|A\rangle$ is written as $\langle A|$, not $\langle A^*|$. That's because the symbol inside the brackets of a ket or a bra is simply a name. For a ket, that name is the name of the vector that the ket represents. But for a bra, the name inside the brackets is the name of the ket to which the bra corresponds. So the bra $\langle A|$ corresponds to the ket $|A\rangle$, but the components of bra $\langle A|$ are the complex conjugates of the components of $|A\rangle$.

You may want to think of a bra like this:


Tells you that this is a device for
turning a vector (ket) into a scalar
< Label |
Name of the vector (ket) to
which this bra corresponds

Main Ideas of This Section

In Dirac notation, a vector is represented as a basis-independent ket, and its components in a specified basis are represented by a column vector. Every ket has a corresponding bra; its components in a specified basis are the complex conjugates of the components of the corresponding ket and are represented by a row vector. The inner product of two vectors is formed by multiplying the bra corresponding to the first vector by the ket corresponding to the second vector, making a "bra-ket" or "bracket."

Relevance to Quantum Mechanics

The solutions to the Schrödinger equation are functions of space and time called quantum wavefunctions, which are the projections of **quantum states** onto a specified basis system. Quantum states may be usefully represented as kets in quantum mechanics. As kets, quantum states are not tied to any particular basis system, but they may be expanded using basis states of position, momentum, energy, or other quantities. Dirac notation is also helpful in providing basis-independent representation of inner products, Hermitian operators (Section 2.3), projection operators (Section 2.4), and expectation values (Section 2.5).

1.3 Abstract Vectors and Functions

To understand the use of bras and kets in quantum mechanics, it's necessary to generalize the concepts of vector components and basis vectors to functions. I think the best way to do that is to change the way you graph vectors. Instead of attempting to replicate three-dimensional physical space as in Fig. 1.4a, simply line up the vector components along the horizontal axis of a two-dimensional graph, with the vertical axis representing the amplitude of the components, as in Fig. 1.4b.

At first glance, a two-dimensional graph of vector components may seem less useful than a three-dimensional graph, but its value becomes clear when you consider spaces with more than three dimensions.

And why would you want to do that? Because higher-dimensional **abstract spaces** turn out to be very useful tools for solving problems in several areas of physics, including classical and quantum mechanics. These spaces are called "abstract" because they're nonphysical – that is, their dimensions don't represent the physical dimensions of the universe we inhabit. For example, an abstract space might consist of all of the values of the parameters of a mathematical model, or all of the possible configurations of a system. So the axes could represent speed, momentum, acceleration, energy, or any other parameter of interest.

Now imagine drawing a set of axes in an abstract space and marking each axis with the values of a parameter. That makes each parameter a "generalized coordinate"; "generalized" because these are not spatial coordinates (such as x, y, and z), but a "coordinate" nonetheless because each location on the axis

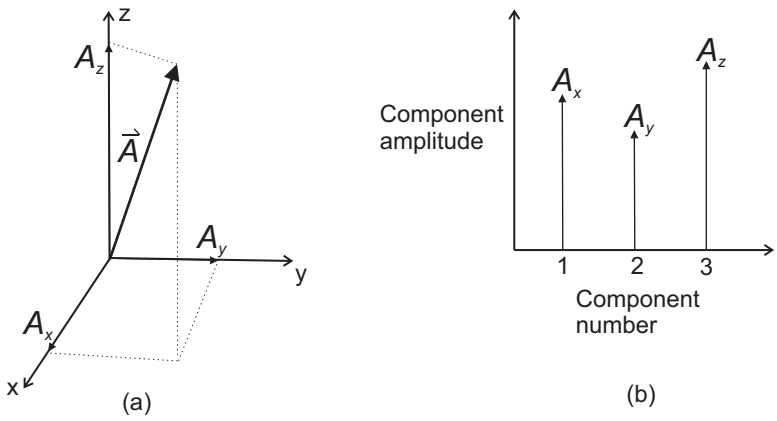


Figure 1.4 Vector components graphed in (a) 3-D and (b) 2-D.

represents a position in the abstract space. So if speed is used as a generalized coordinate, an axis might represent the range of speeds from 0 to 20 meters per second, and the "distance" between two points on that axis is simply the difference between the speeds at those two points.

Physicists sometimes refer to "length" and "direction" in an abstract space, but you should remember that in such cases "length" is not a physical distance, but rather the difference in coordinate values at two locations. And "direction" is not a spatial direction, but rather an angle relative to an axis along which a parameter changes.

The multidimensional space most useful in quantum mechanics is an abstract vector space called "**Hilbert space**," after the German mathematician David Hilbert. If this is your first encounter with Hilbert space, don't panic. You'll find all the basics you need to understand the vector space of quantum wavefunctions in this book, and most comprehensive texts on quantum mechanics such as those in the Bibliography provide additional details, if you want greater depth.

To understand the characteristics of Hilbert space, recall that vector spaces are collections of vectors that behave according to certain rules, such as vector addition and scalar multiplication. In addition to those rules, an "inner product space" also includes rules for multiplying two vectors together (the generalized scalar product). But an issue arises when forming the inner product between two higher-dimensional vectors, and to understand that issue, consider the graph of the components of an N-dimensional vector shown in Fig. 1.5.

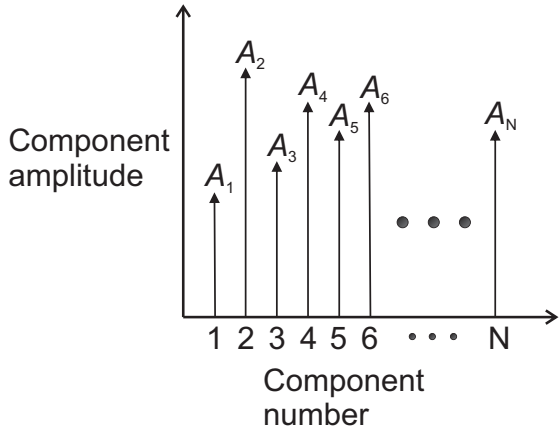


Figure 1.5 Vector components of an N-dimensional vector.

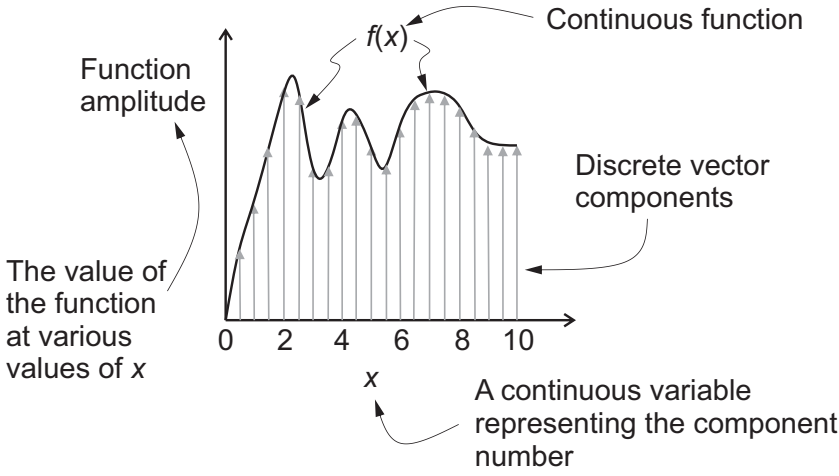

Continuous function
f(x)
Function amplitude
Discrete vector components
The value of the function at various values of x
0 2 4 6 8 10
x
A continuous variable representing the component number

Figure 1.6 Relationship between vector components and continuous function.

Just as each of the three components (A_x, A_y, and A_z) pertains to a basis vector ($\hat{\imath}$, $\hat{\jmath}$, and $\hat{k}$), each of the N components in Fig. 1.5 pertains to a basis vector in the N-dimensional abstract vector space inhabited by the vector.

Now imagine how such a graph would appear for a vector with an even larger number of components. The more components that you display on your graph for a given range, the closer together those components will appear along the horizontal axis, as shown in Fig. 1.6. If you're dealing with a vector with an extremely large number of components, the components may be treated as a continuous function rather than a set of discrete values. That function (call it "f") is depicted as the curvy line connecting the tips of the vector components in Fig. 1.6. As you can see, the horizontal axis is labeled with a continuous variable (call it "x"), which means that the amplitudes of the components are represented by the continuous function $f(x)$.[4]

So the continuous function $f(x)$ is composed of a series of amplitudes, with each amplitude pertaining to a different value of the continuous variable x. And a vector is composed of a series of component amplitudes, with each component pertaining to a different basis vector.

In light of this parallel between a continuous function such as $f(x)$ and the components of a vector such as $\vec{A}$, it's probably not surprising that the rules for

[4] We're dealing with functions of a single variable called x, but the same concepts apply to functions of multiple variables.

addition and scalar multiplication apply to functions as well as vectors. So two functions $f(x)$ and $g(x)$ add to produce a new function, and that addition is done by adding the value of $f(x)$ to the value of $g(x)$ at every x (just as the addition of two vectors is done by adding corresponding components for each basis vector). Likewise, multiplying a function by a scalar results in a new function, which has a value at every x of the original function $f(x)$ times the scalar multiplier (just as multiplying a vector by a scalar produces a new vector with each component amplitude multiplied by the scalar).

But what about the inner product? Is there an equivalent process for continuous functions? Yes, there is. Since you know that for vectors the dot product in an orthonormal system can be found by summing the products of corresponding components in a given basis (such as $A_x B_x + A_y B_y + A_z B_z$), a reasonable guess is that the equivalent operation for continuous functions such as $f(x)$ and $g(x)$ involves multiplication of the functions followed by integration rather than discrete summation. That works – the inner product between two functions $f(x)$ and $g(x)$ (which, like vectors, may be represented by kets) is found by integrating their product over x:

$$(f(x), g(x)) = \langle f(x)|g(x)\rangle = \int_{-\infty}^{\infty} f^*(x)g(x)dx, \tag{1.19}$$

in which the asterisk after the function $f(x)$ in the integral represents the complex conjugate, as in Eq. 1.16. The reason for taking the complex conjugate is explained in the next section.

And what's the significance of the inner product between two functions? Recall that the dot product between two vectors uses the projection of one vector onto the direction of the other to tell you how much one vector "lies along" the direction of the other. Similarly, the inner product between two functions uses the "projection" of one function onto the other to tell you how much of one function "lies along" the other (or, if you prefer, how much one function gets you in the "direction" of the other function).[5]

Obeying the rules for addition, scalar multiplication, and the inner product means that functions like $f(x)$ can behave like vectors – they are not members of the vector space of three-dimensional physical vectors, but they are members of their own abstract vector space.

There is, however, one more condition that must be satisfied before we can call that vector space a Hilbert space. That condition is that the functions must have a finite norm:

[5] The concept of the "direction" of a function may make more sense after you've read about orthogonal functions in Section 1.5.

$$|f(x)|^2 = \langle f(x)|f(x)\rangle = \int_{-\infty}^{\infty} f^*(x)f(x)dx < \infty. \tag{1.20}$$

In other words, the integral of the square of every function in this space must converge to a finite value. Such functions are said to be "square summable" or "square integrable."

Main Ideas of This Section
Real vectors in physical 3D space have length and direction, and abstract vectors in higher-dimensional space have generalized "length" (determined by their norm) and "direction" (determined by their projection onto other vectors). Just as a vector is composed of a series of component amplitudes, each pertaining to a different basis vector, a continuous function is composed of a series of amplitudes, each pertaining to a different value of a continuous variable. These continuous functions have generalized "length" and "direction" and obey the rules of vector addition, scalar multiplication, and the inner product. Hilbert space is a collection of such functions that also have finite norm.

Relevance to Quantum Mechanics
The solutions to the Schrödinger equation are quantum wavefunctions that may be treated as abstract vectors. This means that concepts such as basis functions, components, orthogonality, and the inner product as a projection along the "direction" of another function may be employed in the analysis of quantum wavefunctions. As you'll see in Chapter 4, these wavefunctions represent probability amplitudes, and the integral of the square of these amplitudes must remain finite to keep the probability finite. So to be physically realizable, quantum wavefunctions must be "normalizable" by dividing by their norms, and their norms must be finite. Hence quantum wavefunctions reside in Hilbert space.

1.4 Complex Numbers, Vectors, and Functions

The motivation for the sequence of Figs. 1.4, 1.5, and 1.6 is to help you understand the relationship between vectors and functions, and that understanding will be very helpful when you're analyzing the solutions to the Schrödinger equation. But as you'll see in Chapter 3, one important difference between

the Schrödinger equation and the classical wave equation is the presence of the **imaginary unit** "i" (the square root of minus one), which means that the wavefunction solutions to the Schrödinger equation may be complex.[6] So this section contains a short review of **complex numbers** and their use in the context of vector components and Dirac notation.

As mentioned in the previous section, the process of taking an inner product between vectors or functions is slightly different for complex quantities. How can a vector be complex? By having complex components. To see why that has an effect on the inner product, consider the length of a vector with complex components. Remember, complex quantities can be purely real, purely imaginary, or a mixture of real and imaginary parts. So the most general way of representing a complex quantity z is

$$z = x + iy, \tag{1.21}$$

in which x is the real part of z and y is the imaginary part of z (be sure not to confuse the imaginary unit $i = \sqrt{-1}$ in this equation with the $\hat{\imath}$ unit vector – you can always tell the difference by noting the caret hat on the unit vector $\hat{\imath}$).

Imaginary numbers are every bit as "real" as real numbers, but they lie along a different number line. That number line is perpendicular to the real number line, and a two-dimensional plot of both number lines represents the "complex plane" shown in Fig. 1.7.

As you can see from this figure, knowing the real and imaginary parts of a complex number allows you to find the magnitude or norm of that number. The magnitude of a complex number is the distance between the point representing the complex number and the origin in the complex plane, and you can find that distance using the **Pythagorean theorem**

$$|z|^2 = x^2 + y^2. \tag{1.22}$$

But if you try to square the complex number z by multiplying by itself, you find

$$z^2 = z \times z = (x + iy) \times (x + iy) = x^2 + 2ixy - y^2, \tag{1.23}$$

which is a complex number, and which may be negative. But a distance should be a real and positive number, so this is clearly not the way to find the distance of z from the origin.

[6] Mathematicians say that such functions are members of an abstract linear vector space "over the field of complex numbers." That means that the components may be complex, and that the rules for scaling a function by multiplying by a scalar apply not only to real scalars, but complex numbers as well.

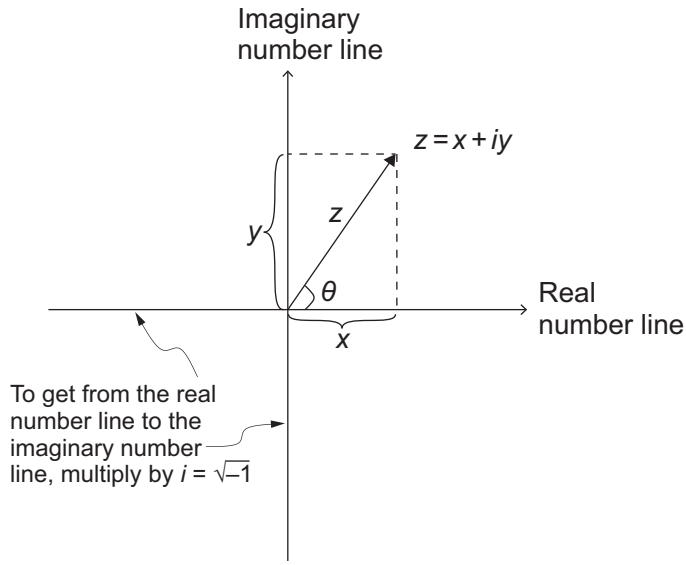


Figure 1.7 Complex number $z = x + iy$ in the complex plane.

To correctly find the magnitude of a complex quantity, it's necessary to multiply the quantity not by itself, but by its complex conjugate. To take the complex conjugate of a complex number, just change the sign of the imaginary part of the number. The complex conjugate is usually indicated by an asterisk, so for the complex quantity $z = x + iy$, the complex conjugate is

$$z^* = x - iy. \tag{1.24}$$

Multiplying by the complex conjugate ensures that the magnitude of a complex number will be real and positive (as long as the real and the imaginary parts are not both zero). You can see that by writing out the terms of the multiplication:

$$|z|^2 = z \times z^* = (x + iy) \times (x - iy) = x^2 - xiy + iyx + y^2 = x^2 + y^2, \tag{1.25}$$

as expected. And since the magnitude (or norm) of a vector $\vec{A}$ can be found by taking the square root of the inner product of the vector with itself, the complex conjugate is built into the process of taking the inner product between complex quantities:

$$|A| = \sqrt{\vec{A} \circ \vec{A}} = \sqrt{A_x^* A_x + A_y^* A_y + A_z^* A_z} = \sqrt{\sum_{i=1}^{N} A_i^* A_i}. \tag{1.26}$$

This also applies to complex functions:

$$|f(x)| = \sqrt{\langle f(x)|f(x)\rangle} = \sqrt{\int_{-\infty}^{\infty} f^*(x)f(x)dx}. \tag{1.27}$$

So it's necessary to use the complex conjugate to find the norm of a complex vector or function. If the inner product involves two different vectors or functions, by convention the complex conjugate is taken of the *first* member of the pair:

$$\begin{aligned} \vec{A} \circ \vec{B} &= \sum_{i=1}^{N} A_i^* B_i \\ \langle f(x)|g(x)\rangle &= \int_{-\infty}^{\infty} f^*(x)g(x)dx. \end{aligned} \tag{1.28}$$

This is the reason for the complex conjugation in the earlier discussion of the inner product using bras and kets (Eqs. 1.16 and 1.19).

The requirement to take the complex conjugate of one member of the inner product for complex vectors and functions means that the order matters, so $\vec{A} \circ \vec{B}$ is not the same as $\vec{B} \circ \vec{A}$. That's because

$$\begin{aligned} \vec{A} \circ \vec{B} &= \sum_{i=1}^{N} A_i^* B_i = \sum_{i=1}^{N} \left(A_i B_i^*\right)^* = \sum_{i=1}^{N} (B_i^* A_i)^* = (\vec{B} \circ \vec{A})^* \\ \langle f(x)|g(x)\rangle &= \int_{-\infty}^{\infty} f^*(x)g(x)dx = \int_{-\infty}^{\infty} [g^*(x)f(x)]^* dx = (\langle g(x)|f(x)\rangle)^*. \end{aligned} \tag{1.29}$$

So reversing the order of the complex vectors or functions in an inner product produces a result that is the complex conjugate of the inner product without switching.

The convention of applying the complex conjugate to the first member of the inner product is common but not universal in physics texts, so you should be aware that you may find some texts and online resources that apply the complex conjugate to the second member.

Main Idea of This Section
Abstract vectors may have complex components, and continuous functions may have complex values. When an inner product is taken between two such vectors or functions, the complex conjugate of the first member must be taken before the product is formed. This ensures that taking the inner product of a complex vector or function with itself produces a real, positive scalar, as required for the norm.

Relevance to Quantum Mechanics
Solutions to the Schrödinger equation may be complex, so when finding the norm of such functions or when taking the inner product between two such functions, it's necessary to take the complex conjugate of the first member of the inner product.

Before moving on to operators and eigenvalues in Chapter 2, you should make sure you have a firm understanding of the meaning of orthogonality of functions and the use of the inner product to find the components of complex vectors and functions. Those are the subjects of the next two sections.

1.5 Orthogonal Functions

For vectors, the concept of orthogonality is straightforward: two vectors are orthogonal if their scalar product is zero, which means that the projection of one of the vectors onto the direction of the other has zero length. Simply put, orthogonal vectors lie along perpendicular lines, as shown in Fig. 1.8a for the two-dimensional vectors $\vec{A}$ and $\vec{B}$ (which we'll take as real for simplicity).

Now consider the plots of the Cartesian components of vectors $\vec{A}$ and $\vec{B}$ in Fig. 1.8b. You can learn something about the relationship between these components by writing out the scalar product of $\vec{A}$ and $\vec{B}$:

$$\begin{aligned} \vec{A} \circ \vec{B} = A_x B_x + A_y B_y &= 0 \\ A_x B_x &= -A_y B_y \\ \frac{A_x}{A_y} &= -\frac{B_y}{B_x}. \end{aligned}$$

This can only be true if one (and only one) of the components of $\vec{A}$ has the opposite sign of the corresponding component of $\vec{B}$. In this case, since $\vec{A}$ points up and to the right (that is, A_x and A_y are both positive), to be perpendicular, $\vec{B}$ must point either up and to the left (with B_x negative and B_y positive, as shown in Fig. 1.8a), or down and to the right (with B_x positive and B_y negative). Additionally, since the angle between the x- and y-axes is 90°, if $\vec{A}$ and $\vec{B}$ are perpendicular, the angle between $\vec{A}$ and the positive x-axis (shown as θ in Fig. 1.8a) must be the same as the angle between $\vec{B}$ and the positive y-axis (or negative y-axis had we taken the "down and to the right" option for $\vec{B}$). For those angles to be the same, the ratio of $\vec{A}$'s components (A_x/A_y) must

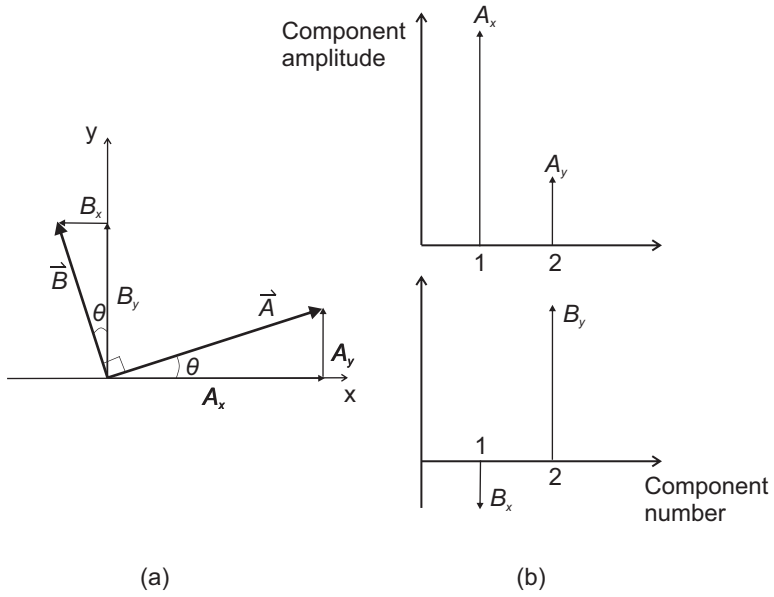


Figure 1.8 (a) Conventional graph of vectors and showing Cartesian components and (b) 2-D graphs of component amplitude vs. component number.

have the same magnitude as the inverse ratio of $\vec{B}$'s components (B_y/B_x). You can get an idea of this inverse ratio in Fig. 1.8b.

Similar considerations apply to N-dimensional abstract vectors as well as continuous functions, as shown in Fig. 1.9a and b.[7] If the N-dimensional abstract vectors $\vec{A}$ and $\vec{B}$ in Fig. 1.9a (again taken as real) are orthogonal, then their inner product $(\vec{A}, \vec{B})$ must equal zero:

$$(\vec{A}, \vec{B}) = \sum_{i=1}^{N} A_i^* B_i = A_1 B_1 + A_2 B_2 + \cdots + A_N B_N = 0.$$

For this sum to be zero, it must be true that some of the component products have opposite signs of others, and the total of all the negative products must equal the total of all the positive products. In the case of the two N-dimensional vectors shown in Fig. 1.9a, the components in the left half of $\vec{B}$ have the same sign as the corresponding components of $\vec{A}$, so the products of those left-half components ($A_i B_i$) are all positive. But the components in the right half of $\vec{B}$

[7] The amplitudes of these components are taken to be sinusoidal in anticipation of the Fourier theory discussion of Section 4.4.

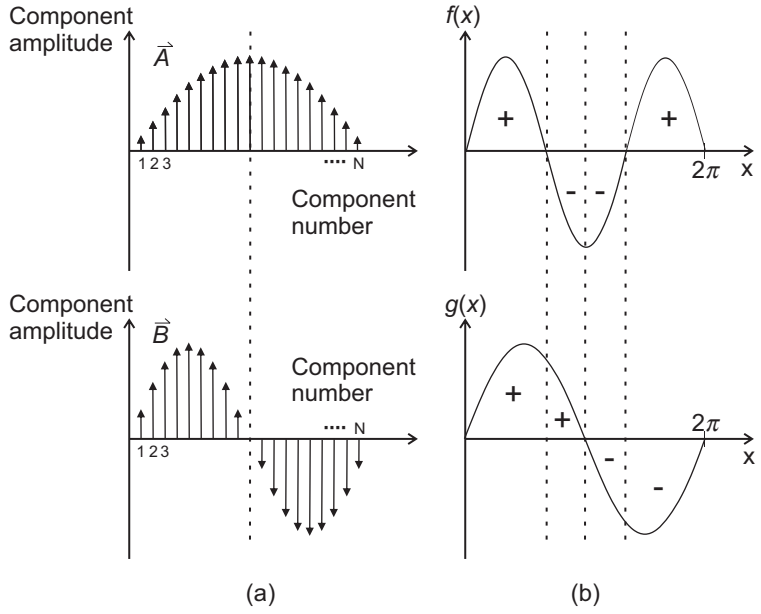


Figure 1.9 Orthogonal N-dimensional vectors (a) and functions (b).

have the opposite sign of the corresponding components in $\vec{A}$, so those products are all negative.

Since the magnitudes of these two vectors are symmetric about their midpoints, the magnitude of the sum of the left-half products equals the magnitude of the sum of the right-half products. With equal magnitudes and opposite signs, the sum of the products of the components from the left half and the right half is zero.

So although $\vec{A}$ and $\vec{B}$ are abstract vectors with "directions" only with respect to generalized rather than spatial coordinates, these two N-dimensional vectors satisfy the requirements of orthogonality, just as the two spatial vectors did in Fig. 1.8. Stated another way, even though we have no way of drawing the N dimensions of these vectors in different physical directions in our three-dimensional space, the zero inner product of $\vec{A}$ and $\vec{B}$ means that the projection of vector $\vec{A}$ onto vector $\vec{B}$ (and of $\vec{B}$ onto $\vec{A}$) has zero "length" in their N-dimensional vector space.

By this point, you've probably realized how orthogonality applies to functions such as $f(x)$ and $g(x)$, shown in Fig. 1.9b. Since these functions

(also taken as real for simplicity) are continuous, the inner-product sum becomes an integral, as described in the previous section. For these functions, the statement of orthogonality is

$$(f(x), g(x)) = \langle f(x)|g(x)\rangle = \int_{-\infty}^{\infty} f^*(x)g(x)dx = \int_{-\infty}^{\infty} f(x)g(x)dx = 0.$$

Just as in the case of discrete vectors $\vec{A}$ and $\vec{B}$, the product $f(x)g(x)$ can be thought of as multiplying the value of the function $f(x)$ by the value of the function $g(x)$ at each value of x. Integrating this product over x is the equivalent of finding the area under the curve formed by the product $f(x)g(x)$.

In the case of the functions $f(x)$ and $g(x)$ in Fig. 1.9b, you can estimate the result of multiplying the two functions and integrating (continuously summing) the result. To do that, notice that for the first one-third of the range of x shown on the graph (left of the first dashed vertical line), $f(x)$ and $g(x)$ have the same sign (both positive). For the next one-sixth of the graph (between the first and second dashed vertical lines), the two functions have opposite signs ($f(x)$ negative and $g(x)$ positive). For the next one-sixth of the graph (between the second and third dashed vertical lines), the signs of $f(x)$ and $g(x)$ are again the same (both negative), and for the final one-third of the graph (right of the third dashed vertical line), the signs are opposite. Due to the symmetry of the regions in which the product $f(x)g(x)$ is positive and negative, the total sum is zero, and these two functions qualify as orthogonal over this range of x. So these two functions are orthogonal in this region in exactly the same way as vectors $\vec{A}$ and $\vec{B}$ are orthogonal.

If you prefer a more mathematical approach to determining the orthogonality of these functions, notice that $g(x)$ is the $\sin x$ function over the range of $x = 0$ to 2π and that $f(x)$ is the $\sin \frac{3}{2}x$ function over the same range. The inner product of these two functions is

$$\begin{aligned}\langle f(x)|g(x)\rangle = \int_{-\infty}^{\infty} f^*(x)g(x)dx &= \int_0^{2\pi} \sin\left(\frac{3}{2}x\right)\sin(x)dx \\ &= \left[\sin\frac{x}{2} - \frac{1}{5}\sin\frac{5x}{2}\right]\Bigg|_0^{2\pi} = 0,\end{aligned}$$

which is consistent with the result obtained by estimating the area under the curve of the product $f(x)g(x)$. You can read more about the orthogonality of harmonic (sine and cosine) functions in Section 4.4.

Main Idea of This Section
Just as the vectors in three-dimensional physical space must be perpendicular if their scalar product is zero, N-dimensional abstract vectors and continuous functions are defined as orthogonal if their inner product is zero.

Relevance to Quantum Mechanics
As you'll see in Section 2.5, orthogonal basis functions play an important role in determining the possible outcomes of measurements of quantum observables and the probability of each outcome.

Orthogonal functions are extremely useful in physics, for reasons that are similar to the reasons that orthogonal coordinate systems are useful. The final section of this chapter shows you how to use the inner product and orthogonal functions to determine the components of multi-dimensional abstract vectors.

1.6 Finding Components Using the Inner Product

As discussed in Section 1.1, the components of a vector that has been expanded using unit vectors (such as $\hat{\imath}$, $\hat{\jmath}$, and $\hat{k}$ in the Cartesian coordinate system) can be written as the scalar product of each unit vector with the vector:

$$\begin{aligned} A_x &= \hat{\imath} \circ \vec{A} \\ A_y &= \hat{\jmath} \circ \vec{A} \\ A_z &= \hat{k} \circ \vec{A}, \end{aligned} \tag{1.30}$$

which can be concisely written as

$$A_i = \hat{\epsilon}_i \circ \vec{A} \qquad i = 1, 2, 3, \tag{1.31}$$

in which $\hat{\epsilon}_1$ represents $\hat{\imath}$, $\hat{\epsilon}_2$ represents $\hat{\jmath}$, and $\hat{\epsilon}_3$ represents $\hat{k}$.

This can be generalized to find the components of an N-dimensional abstract vector represented by the ket $|A\rangle$ in a basis system with orthogonal basis vectors $\vec{\epsilon}_1, \vec{\epsilon}_2, \ldots \vec{\epsilon}_N$:

$$A_i = \frac{\vec{\epsilon}_i \circ \vec{A}}{|\vec{\epsilon}_i|^2} = \frac{\langle \epsilon_i | A \rangle}{\langle \epsilon_i | \epsilon_i \rangle}. \tag{1.32}$$

Notice that the basis vectors in this case are orthogonal, but they don't necessarily have unit length (as you can tell by their hats, which are regular

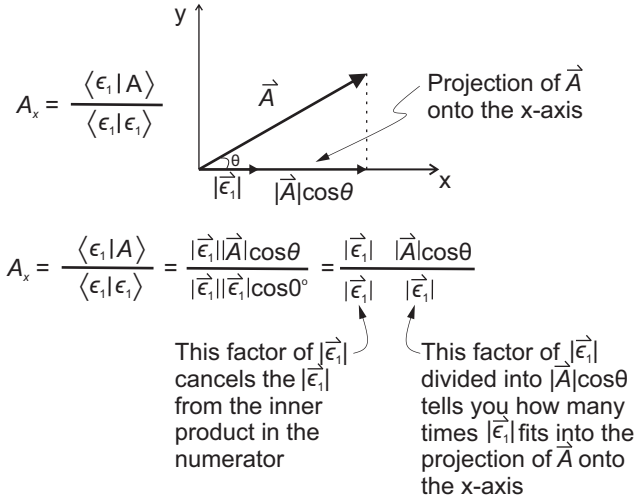

y
x
θ
Projection of $\vec{A}$ onto the x-axis
$A_x = \frac{\langle \epsilon_1 | A \rangle}{\langle \epsilon_1 | \epsilon_1 \rangle}$
$|\vec{\epsilon}_1|$
$|\vec{A}|\cos\theta$
$A_x = \frac{\langle \epsilon_1 | A \rangle}{\langle \epsilon_1 | \epsilon_1 \rangle} = \frac{|\vec{\epsilon}_1||\vec{A}|\cos\theta}{|\vec{\epsilon}_1||\vec{\epsilon}_1|\cos 0^\circ} = \frac{|\vec{\epsilon}_1|}{|\vec{\epsilon}_1|} \frac{|\vec{A}|\cos\theta}{|\vec{\epsilon}_1|}$
This factor of $|\vec{\epsilon}_1|$ cancels the $|\vec{\epsilon}_1|$ from the inner product in the numerator
This factor of $|\vec{\epsilon}_1|$ divided into $|\vec{A}|\cos\theta$ tells you how many times $|\vec{\epsilon}_1|$ fits into the projection of $\vec{A}$ onto the x-axis

Figure 1.10 Normalizing the inner product for a basis vector with non-unit length.

vector hats ($\vec{\ }$) rather than unit-vector hats ($\hat{\ }$). In that case, to find the vector's components using the inner product, it's necessary to divide the result of the inner product by the square of the basis vector's length, as you can see in the denominators of the fractions in Eq. 1.32. This factor wasn't necessary in Eqs. 1.30 or 1.31 because each Cartesian unit vector $\hat{\imath}$, $\hat{\jmath}$, and $\hat{k}$ has a length of one.

If you're wondering why it's necessary to divide by the square rather than the first power of the length of each basis vector, consider the situation shown in Fig. 1.10.

In this figure, basis vector $\vec{\epsilon}_1$ points along the x-axis, and the angle between vector $\vec{A}$ and the positive x-axis is θ. The projection of vector $\vec{A}$ onto the x-axis is $|\vec{A}|\cos\theta$; Eq. 1.32 gives the x-component of $\vec{A}$ as

$$A_x = \frac{\vec{\epsilon}_1 \circ \vec{A}}{|\vec{\epsilon}_1|^2} = \frac{\langle \epsilon_1 | A \rangle}{\langle \epsilon_1 | \epsilon_1 \rangle}. \tag{1.33}$$

As shown in Fig. 1.10, the two factors of $|\vec{\epsilon}_1|$ in the denominator of Eq. 1.33 are exactly what's needed to give A_x in units of $|\vec{\epsilon}_1|$. That's because one factor of $|\vec{\epsilon}_1|$ cancels the same factor from the inner product in the numerator, and the second factor of $|\vec{\epsilon}_1|$ converts $|\vec{A}|\cos\theta$ into the number of "steps" of $|\vec{\epsilon}_1|$ that fit into the projection of $\vec{A}$ onto the x-axis.

So if, for example, vector $\vec{A}$ is a real spatial vector with length $|\vec{A}|$ of 10 km at an angle of 35° to the x-axis, then the projection of $\vec{A}$ onto the x-axis ($|\vec{A}|\cos\theta$) is about 8.2 km. But if the basis vector $\vec{\epsilon}_1$ has length of 2 km, dividing 8.2 km by 2 km gives 4.1 "steps" of 2 km, so the x-component of $\vec{A}$ is $A_x = 4.1$ (not 4.1 km, because the units are carried by the basis vectors).

Had you chosen a basis vector with length of one unit (of the units in which vector $\vec{A}$ is measured, which is kilometers in this example), then the denominator of Eq. 1.33 would have a value of one, and the number of steps along the x-axis would be 8.2.

The process of dividing by the square of the norm of a vector or function is called "**normalization,**" and orthogonal vectors or functions that have a length of one unit are "orthonormal." The condition of orthonormality for basis vectors $\vec{\epsilon}$ is often written as

$$\vec{\epsilon}_i \circ \vec{\epsilon}_j = \langle \epsilon_i | \epsilon_j \rangle = \delta_{i,j}, \tag{1.34}$$

in which $\delta_{i,j}$ represents the **Kronecker delta**, which has a value of one if $i = j$ or zero if $i \neq j$.

The expansion of a vector as the weighted combination of a set of basis vectors and the use of the normalized scalar product to find the vector's components for a specified basis can be extended to the functions of Hilbert space. Expressing these functions as kets, the expansion of function $|\psi\rangle$ using basis functions $|\psi_n\rangle$ is

$$|\psi\rangle = c_1 |\psi_1\rangle + c_2 |\psi_2\rangle + \cdots + c_N |\psi_N\rangle = \sum_{n=1}^{N} c_n |\psi_n\rangle, \tag{1.35}$$

in which c_1 tells you the "amount" of basis function $|\psi_1\rangle$ in function $|\psi\rangle$, c_2 tells you the "amount" of basis function $|\psi_2\rangle$ in function $|\psi\rangle$, and so on. As long as the basis functions $|\psi_1\rangle$, $|\psi_2\rangle \ldots |\psi_N\rangle$ are orthogonal, the components $c_1, c_2, \ldots c_N$ can be found using the normalized inner product:

$$\begin{aligned} c_1 &= \frac{\langle \psi_1 | \psi \rangle}{\langle \psi_1 | \psi_1 \rangle} = \frac{\int_{-\infty}^{\infty} \psi_1^*(x)\psi(x)dx}{\int_{-\infty}^{\infty} \psi_1^*(x)\psi_1(x)dx} \\ c_2 &= \frac{\langle \psi_2 | \psi \rangle}{\langle \psi_2 | \psi_2 \rangle} = \frac{\int_{-\infty}^{\infty} \psi_2^*(x)\psi(x)dx}{\int_{-\infty}^{\infty} \psi_2^*(x)\psi_2(x)dx} \\ c_N &= \frac{\langle \psi_N | \psi \rangle}{\langle \psi_N | \psi_N \rangle} = \frac{\int_{-\infty}^{\infty} \psi_N^*(x)\psi(x)dx}{\int_{-\infty}^{\infty} \psi_N^*(x)\psi_N(x)dx}, \end{aligned} \tag{1.36}$$

in which each numerator represents the projection of function $|\psi\rangle$ onto one of the basis functions, and each denominator represents the square of the norm of that basis function.

This approach to finding the components of a function (using sinusoidal basis functions) was pioneered by the French mathematician and physicist Jean-Baptiste Joseph Fourier in the early part of the nineteenth century. Fourier theory comprehends both Fourier synthesis, in which periodic functions are synthesized by weighted combination of sinusoidal functions, and Fourier analysis, in which the sinusoidal components of a periodic function are determined using the approach described earlier. In quantum mechanics texts, this process is sometimes called "spectral decomposition," since the weighting coefficients (c_n) are called the "spectrum" of a function.

To see how this works, consider a function $|\psi(x)\rangle$ expanded using the basis functions $|\psi_1\rangle = \sin x$, $|\psi_2\rangle = \cos x$, and $|\psi_3\rangle = \sin 2x$ over the interval $x = -\pi$ to $x = \pi$:

$$\psi(x) = 5\,|\psi_1\rangle - 10\,|\psi_2\rangle + 4\,|\psi_3\rangle\,.$$

In this case, you can read the components $c_1 = 5$, $c_2 = -10$, and $c_3 = 4$ directly from this equation for $\psi(x)$. But to understand how Eq. 1.36 gives these values, write

$$c_1 = \frac{\int_{-\infty}^{\infty} \psi_1^*(x)\psi(x)dx}{\int_{-\infty}^{\infty} \psi_1^*(x)\psi_1(x)dx} = \frac{\int_{-\pi}^{\pi} [\sin x]^* [5\sin x - 10\cos x + 4\sin 2x]\,dx}{\int_{-\pi}^{\pi} [\sin x]^* \sin x\,dx}$$

$$c_2 = \frac{\int_{-\infty}^{\infty} \psi_2^*(x)\psi(x)dx}{\int_{-\infty}^{\infty} \psi_2^*(x)\psi_2(x)dx} = \frac{\int_{-\pi}^{\pi} [\cos x]^* [5\sin x - 10\cos x + 4\sin 2x]\,dx}{\int_{-\pi}^{\pi} [\cos x]^* \cos x\,dx}$$

$$c_3 = \frac{\int_{-\infty}^{\infty} \psi_3^*(x)\psi(x)dx}{\int_{-\infty}^{\infty} \psi_3^*(x)\psi_3(x)dx} = \frac{\int_{-\pi}^{\pi} [\sin 2x]^* [5\sin x - 10\cos x + 4\sin 2x]\,dx}{\int_{-\pi}^{\pi} [\sin 2x]^* \sin 2x\,dx}.$$

These integrals can be evaluated with the help of the relations

$$\int_{-\pi}^{\pi} \sin^2 ax\,dx = \left[\frac{x}{2} - \frac{\sin 2ax}{4a}\right]\Bigg|_{-\pi}^{\pi} = \pi$$

$$\int_{-\pi}^{\pi} \cos^2 ax\,dx = \left[\frac{x}{2} + \frac{\sin 2ax}{4a}\right]\Bigg|_{-\pi}^{\pi} = \pi$$

$$\int_{-\pi}^{\pi} \sin x \cos x\,dx = \left[\frac{1}{2}\sin^2 x\right]\Bigg|_{-\pi}^{\pi} = 0$$

$$\int_{-\pi}^{\pi} \sin mx \sin nx\,dx = \left[\frac{\sin(m-n)x}{2(m-n)} + \frac{\sin(m+n)x}{2(m+n)}\right]\Bigg|_{-\pi}^{\pi} = 0,$$

in which m and n are (different) integers. Applying these gives

$$c_1 = \frac{5(\pi) - 10(0) + 4(0)}{\pi} = 5$$
$$c_2 = \frac{5(0) - 10(\pi) + 4(0)}{\pi} = -10$$
$$c_3 = \frac{5(0) - 10(0) + 4(\pi)}{\pi} = 4,$$

as expected. Notice that in this example the basis functions $\sin x$, $\cos x$, and $\sin 2x$ are orthogonal but not orthonormal, since their norms are π rather than one. Some students express surprise that sinusoidal functions are not normalized, since their values run from -1 to $+1$. But remember that it's the integral of the square of the function, not the peak value, that determines the function's norm.

Once you feel confident in your understanding of functions as members of an abstract vector space, the expansion of vectors and functions using components in a specified basis, Dirac's bra/ket notation, and the role of the inner product in determining the components of vectors and functions, you should be ready to tackle the subjects of operators and eigenfunctions. You can read about those topics in the next chapter, but if you'd like to make sure that you're able to put the concepts and mathematical techniques covered in this chapter into practice before proceeding, you may find the problems in the next section helpful (and if you get stuck or just want to check your solutions, remember that full interactive solutions to every problem are available on the book's website).

1.7 Problems

1. Find the components of vector $\vec{C} = \vec{A} + \vec{B}$ if $\vec{A} = 3\hat{\imath} - 2\hat{\jmath}$ and $\vec{B} = \hat{\imath} + \hat{\jmath}$ using Eq. 1.4. Verify your answer using graphical addition.
2. What are the lengths of vectors $\vec{A}$, $\vec{B}$, and $\vec{C}$ from Problem 1? Verify your answers using your graph from Problem 1.
3. Find the scalar product $\vec{A} \circ \vec{B}$ for vectors $\vec{A}$ and $\vec{B}$ from Problem 1. Use your result to find the angle between $\vec{A}$ and $\vec{B}$ using Eq. 1.10 and the magnitudes $|\vec{A}|$ and $|\vec{B}|$ that you found in Problem 2. Verify your answer for the angle using your graph from Problem 1.
4. Are the 2D vectors $\vec{A}$ and $\vec{B}$ from Problem 1 orthogonal? Consider what happens if you add a third component of $+\hat{k}$ to $\vec{A}$ and $-\hat{k}$ to $\vec{B}$; are the

3D vectors $\vec{A} = 3\hat{\imath} - 2\hat{\jmath} + \hat{k}$ and $\vec{B} = \hat{\imath} + \hat{\jmath} - \hat{k}$ orthogonal? This illustrates the principal that vectors (and abstract N-dimensional vectors) may be orthogonal over some range of components but non-orthogonal over a different range.

5. If ket $|\psi\rangle = 4\,|\epsilon_1\rangle - 2i\,|\epsilon_2\rangle + i\,|\epsilon_3\rangle$ in a coordinate system with orthonormal basis kets $|\epsilon_1\rangle$, $|\epsilon_2\rangle$, and $|\epsilon_3\rangle$, find the norm of $|\psi\rangle$. Then "normalize" $|\psi\rangle$ by dividing each component of $|\psi\rangle$ by the norm of $|\psi\rangle$.

6. For ket $|\psi\rangle$ from Problem 5 and ket $|\phi\rangle = 3i\,|\epsilon_1\rangle + |\epsilon_2\rangle - 5i\,|\epsilon_3\rangle$, find the inner product $\langle\phi|\psi\rangle$ and show that $\langle\phi|\psi\rangle = \langle\psi|\phi\rangle^*$.

7. If m and n are different positive integers, are the functions $\sin mx$ and $\sin nx$ orthogonal over the interval $x = 0$ to $x = 2\pi$? What about over the interval $x = 0$ to $x = \frac{3\pi}{2}$?

8. Can the functions $e^{i\omega t}$ and $e^{2i\omega t}$ with $\omega = \frac{2\pi}{T}$ form an orthonormal basis over the interval $t = 0$ to $t = T$?

9. Given the basis vectors $\vec{\epsilon}_1 = 3\hat{\imath}$, $\vec{\epsilon}_2 = 4\hat{\jmath} + 4\hat{k}$, and $\vec{\epsilon}_3 = -2\hat{\jmath} + \hat{k}$, what are the components of vector $\vec{A} = 6\hat{\imath} + 6\hat{\jmath} + 6\hat{k}$ along the direction of each of these basis vectors?

10. Given square-pulse function $f(x) = 1$ for $0 \le x \le L$ and $f(x) = 0$ for $x < 0$ and $x > L$, find the values of c_1, c_2, c_3, and c_4 for the basis functions $\psi_1 = \sin(\frac{\pi x}{L})$, $\psi_2 = \cos(\frac{\pi x}{L})$, $\psi_3 = \sin(\frac{2\pi x}{L})$, and $\psi_4 = \cos(\frac{2\pi x}{L})$ over the same interval.

2

Operators and Eigenfunctions

The concepts and techniques discussed in the previous chapter are intended to prepare you to cross the bridge between the mathematics of vectors and functions and the expected results of measurements of quantum observables such as position, momentum, and energy. In quantum mechanics, every physical observable is associated with a linear "operator" that can be used to determine possible measurement outcomes and their probabilities for a given quantum state.

This chapter begins with an introduction to operators, eigenvectors, and eigenfunctions in Section 2.1, followed by an explanation of the use of Dirac notation with operators in Section 2.2. Hermitian operators and their importance are discussed in Section 2.3, and projection operators are introduced in Section 2.4. The calculation of expectation values is the subject of Section 2.5, and as in every chapter, you'll find a series of problems to test your understanding in the final section.

2.1 Operators, Eigenvectors, and Eigenfunctions

If you've heard the phrase "quantum operator" and you're wondering "What exactly is an operator?," you'll be happy to learn that an operator is simply an instruction to perform a certain process on a number, vector, or function. You've undoubtedly seen operators before, although you may not have called them that. But you know that the symbol "$\sqrt{\ }$" is an instruction to take the square root of whatever appears under the roof of the symbol, and "$d(\ )/dx$" tells you to take the first derivative with respect to x of whatever appears inside the parentheses.

The operators you'll encounter in quantum mechanics are called "linear" because applying them to a sum of vectors or functions gives the same result as applying them to the individual vectors or functions and then summing the results. So if $\widehat{O}$ is a linear operator[1] and f_1 and f_2 are functions, then

$$\widehat{O}(f_1 + f_2) = \widehat{O}(f_1) + \widehat{O}(f_2). \tag{2.1}$$

Linear operators also have the property that multiplying a function by a scalar and then applying the operator gives the same result as first applying the operator and then multiplying the result by the scalar. So if c is a (potentially complex) scalar and f is a function, then

$$\widehat{O}(cf) = c\widehat{O}(f), \tag{2.2}$$

if $\widehat{O}$ is a linear operator.

To understand the operators used in quantum mechanics, I think it's helpful to begin by representing an operator as a square matrix and considering what happens when you multiply a matrix and a vector (in quantum mechanics there are times when it's easier to comprehend a process by considering matrix mathematics, and this is one of those times). From the rules of matrix multiplication, you may remember that multiplying a matrix ($\bar{\bar{R}}$) by a column vector $\vec{A}$ works like this[2]:

$$\bar{\bar{R}}\vec{A} = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = \begin{pmatrix} R_{11}A_1 + R_{12}A_2 \\ R_{21}A_1 + R_{22}A_2 \end{pmatrix}. \tag{2.3}$$

This type of multiplication can be done only when the number of columns of the matrix equals the number of rows of the vector (two in this case, since $\vec{A}$ has two components). So the process of multiplying a matrix by a vector produces another vector – the matrix has "operated" on the vector, transforming it into another vector. That's why you'll see linear operators described as "linear transformations" in some texts.

What effect does this type of operation have on the vector? That depends on the matrix and on the vector. Consider, for example, the matrix

$$\bar{\bar{R}} = \begin{pmatrix} 4 & -2 \\ -2 & 4 \end{pmatrix}$$

[1] There are several ways of denoting an operator, but the most common in quantum texts is to put a caret hat (ˆ) on top of the operator label.

[2] There doesn't seem to be a standard notation for matrices in quantum books, so I'll use the double-bar hat ($\bar{\bar{}}$) for two-dimensional matrices and the vector symbol ($\vec{}$) or the ket symbol $|\,\rangle$ for single-column matrices.

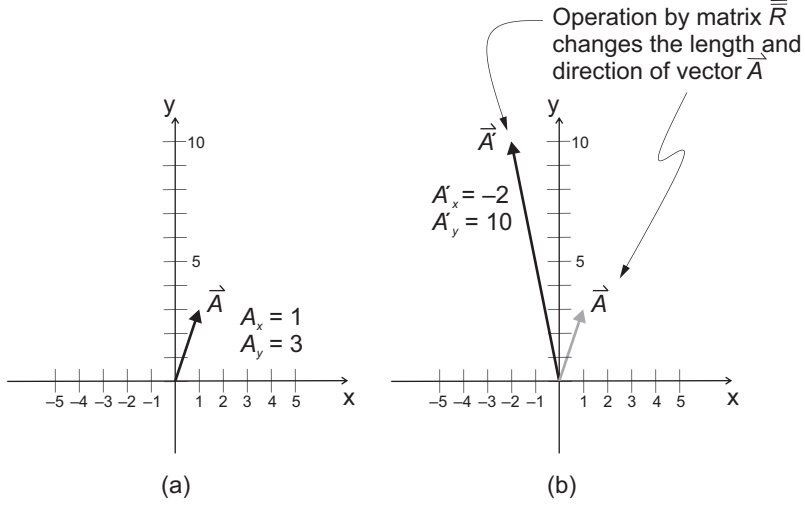


Figure 2.1 Vector $\vec{A}$ before (a) and after (b) operation of matrix $\bar{\bar{R}}$.

and the vector $\vec{A} = \hat{\imath} + 3\hat{\jmath}$, shown in Fig. 2.1a. Writing the components of $\vec{A}$ as a column vector and multiplying gives

$$\bar{\bar{R}}\vec{A} = \begin{pmatrix} 4 & -2 \\ -2 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} (4)(1) + (-2)(3) \\ (-2)(1) + (4)(3) \end{pmatrix} = \begin{pmatrix} -2 \\ 10 \end{pmatrix}. \tag{2.4}$$

So the operation of matrix $\bar{\bar{R}}$ on vector $\vec{A}$ produces another vector that has a different length and points in a different direction. This new vector is shown as vector $\vec{A}'$ in Fig. 2.1b.

Why does a matrix operating on a vector generally change the direction of the vector? You can understand that by realizing that the x-component of the new vector $\vec{A}'$ is a weighted combination of both components of the original vector $\vec{A}$, and the weighting coefficients are provided by the first row of matrix $\bar{\bar{R}}$. Likewise, the y-component of $\vec{A}'$ is a weighted combination of both components of $\vec{A}$, with weighting coefficients provided by the second row of matrix $\bar{\bar{R}}$.

This means that, depending on the values of the matrix elements and the components of the original vector, the weighted combinations will, in general, endow the new vector with a different magnitude from that of the original vector. And here's a key consideration: if the *ratio* of the new components differs from the ratio of the original components, then the new vector will

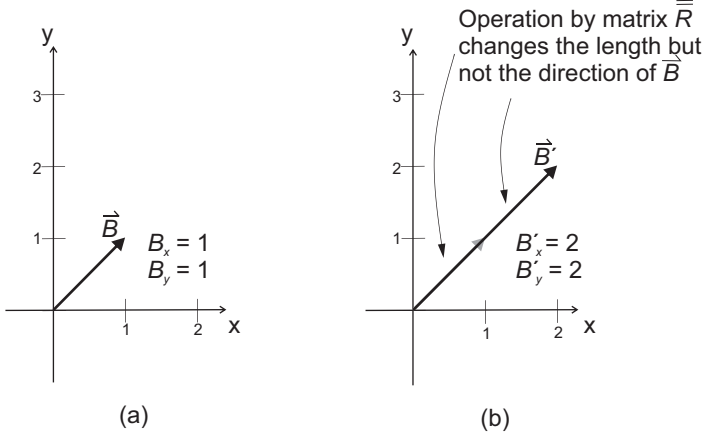


Figure 2.2 Vector $\vec{B}$ before (a) and after (b) operation of matrix $\bar{\bar{R}}$.

point in a different direction from that of the original vector. In such cases, the relative amounts of the basis vectors are changed by the operation of the matrix on the vector.

Now consider the effect of matrix $\bar{\bar{R}}$ on a different vector – for example, vector $\vec{B} = \hat{\imath} + \hat{\jmath}$ shown in Fig. 2.2a. In this case, the multiplication looks like this:

$$\bar{\bar{R}}\vec{B} = \begin{pmatrix} 4 & -2 \\ -2 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} (4)(1) + (-2)(1) \\ (-2)(1) + (4)(1) \end{pmatrix}$$
$$= \begin{pmatrix} 2 \\ 2 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 2\vec{B}. \tag{2.5}$$

So operating on vector $\vec{B}$ with matrix $\bar{\bar{R}}$ stretches the length of $\vec{B}$ to twice its original value but does not change the direction of $\vec{B}$. That means that the relative amounts of the basis vectors in vector $\vec{B}'$ are the same as in vector $\vec{B}$.

A vector for which the direction is not changed after multiplication by a matrix is called an "eigenvector" of that matrix, and the factor by which the length of the vector is scaled is called the "eigenvalue" for that eigenvector (if the vector's length is also unaffected by operation of the matrix, the eigenvalue for that eigenvector equals one). So vector $\vec{B} = \hat{\imath} + \hat{\jmath}$ is an eigenvector of matrix $\bar{\bar{R}}$ with eigenvalue of 2.

Eq. 2.5 is an example of an "eigenvalue equation"; the general form is

$$\bar{\bar{R}}\vec{A} = \lambda\vec{A}, \tag{2.6}$$

in which $\vec{A}$ represents an eigenvector of matrix $\bar{\bar{R}}$ with eigenvalue λ.

The procedure for determining the eigenvalues and eigenvectors of a matrix is not difficult; you can see that procedure and several worked examples on the book's website. If you work through that process for matrix $\bar{\bar{R}}$ in the previous example, you'll find that the vector $\vec{C} = \hat{\imath} - \hat{\jmath}$ is also an eigenvector of matrix $\bar{\bar{R}}$; its eigenvalue is 6.

Here are two helpful hints for the matrices you're likely to encounter in quantum mechanics: the sum of the eigenvalues of a matrix is equal to the trace of the matrix (that is, the sum of the diagonal elements of the matrix, which is 8 in this case), and the product of the eigenvalues is equal to the determinant of the matrix (which is 12 in this case).

Just as matrices act as operators on vectors to produce new vectors, there are mathematical processes that act as operators on functions to produce new functions. If the new function is a scalar multiple of the original function, that function is called an "eigenfunction" of the operator. The eigenfunction equation corresponding to the eigenvector equation (Eq. 2.6) is

$$\widehat{O}\psi = \lambda\psi \tag{2.7}$$

in which ψ represents an eigenfunction of operator $\widehat{O}$ with eigenvalue λ.

You may be wondering what kind of operator works on a function to produce a scaled version of that function. As an example, consider a "derivative operator" $\widehat{D} = \frac{d}{dx}$. To determine whether the function $f(x) = \sin kx$ is an eigenfunction of operator $\widehat{D}$, apply $\widehat{D}$ to $f(x)$ and see if the result is proportional to $f(x)$:

$$\widehat{D}f(x) = \frac{d(\sin kx)}{dx} = k\cos kx \stackrel{?}{=} \lambda(\sin kx). \tag{2.8}$$

So is there any single number (real or complex) that you can multiply by $\sin kx$ to get $k\cos kx$? If you think about the values of $\sin kx$ and $k\cos kx$ at $kx = 0$ and $kx = \pi$ (or look at a graph of these two functions), it should be clear that there's no value of λ that makes Eq. 2.8 true. So $\sin kx$ does not qualify as an eigenfunction of the operator $\widehat{D} = \frac{d}{dx}$.

Now try the same process for the second-derivative operator $\widehat{D^2} = \frac{d^2}{dx^2}$:

$$\widehat{D^2}f(x) = \frac{d^2(\sin kx)}{dx^2} = \frac{d(k\cos kx)}{dx} = -k^2\sin kx \stackrel{?}{=} \lambda(\sin kx). \tag{2.9}$$

In this case, the eigenvalue equation is true if $\lambda = -k^2$. That means that $\sin kx$ is an eigenfunction of the second-derivative operator $\widehat{D^2} = \frac{d^2}{dx^2}$, and the eigenvalue for this eigenfunction is $\lambda = -k^2$.

Main Ideas of This Section
A linear operator may be represented as a matrix that transforms a vector into another vector. If that new vector is a scaled version of the original vector, that vector is an eigenvector of the matrix, and the scaling factor is the eigenvalue for that eigenvector. An operator may also be applied to a function, producing a new function; if that new function is a multiple of the original function, then that function is an eigenfunction of the operator.

Relevance to Quantum Mechanics
In quantum mechanics, every physical observable such as position, momentum, and energy is associated with an operator, and the state of a system may be expressed as a linear combination of the eigenfunctions of that operator. The eigenvalues for those eigenfunctions represent possible outcomes of measurements of that observable.

2.2 Operators in Dirac Notation

To work with quantum-mechanical operators, it's helpful to become familiar with the way operators fit into Dirac notation. Using that notation makes the general eigenvalue equation look like this:

$$\widehat{O}\,|\psi\rangle = \lambda\,|\psi\rangle \tag{2.10}$$

in which ket $|\psi\rangle$ is called an "eigenket" of operator $\widehat{O}$.

Now consider what happens when you form the inner product of ket $|\phi\rangle$ with both sides of this equation:

$$(|\phi\rangle\,, \widehat{O}\,|\psi\rangle) = (|\phi\rangle\,, \lambda\,|\psi\rangle).$$

Remember, in taking an inner product, the first member of the inner product ($|\phi\rangle$ in this case) becomes a bra. Multiplying by the bra $\langle\phi|$ from the left makes this equation

$$\langle\phi|\,\widehat{O}\,|\psi\rangle = \langle\phi|\,\lambda\,|\psi\rangle\,. \tag{2.11}$$

The left side of this equation has an operator "sandwiched" between a bra and a ket, and the right side has a constant in the same position. Expressions like this are extremely common (and useful) in quantum mechanics, so it's worthwhile to spend some time understanding what they mean and how you can use them.

The first thing to realize about an expression such as $\langle\phi|\,\widehat{O}\,|\psi\rangle$ is that it represents a scalar, not a vector or operator. To see why that's true, think about operator $\widehat{O}$ operating to the right on ket $|\psi\rangle$ (you could choose to let it operate to the left on bra $\langle\phi|$, and you'll get the same answer).[3] Just as a matrix operating on a column vector gives another column vector, letting operator $\widehat{O}$ work on ket $|\psi\rangle$ gives another ket, which we'll call $|\psi'\rangle$:

$$\widehat{O}\,|\psi\rangle = |\psi'\rangle. \tag{2.12}$$

This makes the left side of Eq. 2.11

$$\langle\phi|\,\widehat{O}\,|\psi\rangle = \langle\phi|\psi'\rangle. \tag{2.13}$$

This inner product is proportional to the projection of ket $|\psi'\rangle$ onto the direction of ket $|\phi\rangle$, and that projection is a scalar. So sandwiching an operator between a bra and a ket produces a scalar result, but how is that result useful? As you'll see in the final section of this chapter, this type of expression can be used to determine one of the most useful quantities in quantum mechanics. That quantity is the expectation value of measurements of quantum observables.

Before getting to that, here's another way in which an expression of the form of Eq. 2.13 is useful: sandwiching an operator between pairs of basis vectors allows you to determine the elements of the matrix representation of that operator in that basis.

To see how that works, consider an operator $\widehat{A}$, which can be represented as a 2×2 matrix:

$$\bar{\bar{A}} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

in which the elements, A_{11}, A_{12} A_{21}, and A_{22} (collectively referred to as A_{ij}) depend on the basis system, just as the components of a vector depend on the basis vectors to which the components apply.

The matrix elements A_{ij} of the matrix representing operator $\widehat{A}$ in a given basis system can be determined by applying the operator to each of the basis vectors of that system. For example, applying operator $\widehat{A}$ to each of the orthonormal basis vectors $\hat{\epsilon}_1$ and $\hat{\epsilon}_2$ represented by kets $|\epsilon_1\rangle$ and $|\epsilon_2\rangle$, the matrix elements determine the "amount" of each basis vector in the result:

[3] Be careful when using an operator to the left on a bra – this is discussed further in Section 2.3.

$$\begin{aligned}\widehat{A}\,|\epsilon_1\rangle &= A_{11}\,|\epsilon_1\rangle + A_{21}\,|\epsilon_2\rangle \\ \widehat{A}\,|\epsilon_2\rangle &= A_{12}\,|\epsilon_1\rangle + A_{22}\,|\epsilon_2\rangle\,.\end{aligned} \tag{2.14}$$

Notice that it's the columns of $\bar{\bar{A}}$ that determine the amount of each basis vector. Now take the inner product of the first of these equations with the first basis ket $|\epsilon_1\rangle$:

$$\begin{aligned}\langle\epsilon_1|\,\widehat{A}\,|\epsilon_1\rangle &= \langle\epsilon_1|\,A_{11}\,|\epsilon_1\rangle + \langle\epsilon_1|\,A_{21}\,|\epsilon_2\rangle \\ &= A_{11}\,\langle\epsilon_1|\epsilon_1\rangle + A_{21}\,\langle\epsilon_1|\epsilon_2\rangle = A_{11},\end{aligned}$$

since $\langle\epsilon_1|\epsilon_1\rangle = 1$ and $\langle\epsilon_1|\epsilon_2\rangle = 0$ for an orthonormal basis system. Hence the matrix element A_{11} for this basis system can be found using the expression $\langle\epsilon_1|\,\widehat{A}\,|\epsilon_1\rangle$.

Taking the inner product of the second equation in Eq. 2.14 with the first basis ket $|\epsilon_1\rangle$ gives

$$\begin{aligned}\langle\epsilon_1|\,\widehat{A}\,|\epsilon_2\rangle &= \langle\epsilon_1|\,A_{12}\,|\epsilon_1\rangle + \langle\epsilon_1|\,A_{22}\,|\epsilon_2\rangle \\ &= A_{12}\,\langle\epsilon_1|\epsilon_1\rangle + A_{22}\,\langle\epsilon_1|\epsilon_2\rangle = A_{12},\end{aligned}$$

so the matrix element A_{12} for this basis system can be found using the expression $\langle\epsilon_1|\,\widehat{A}\,|\epsilon_2\rangle$.

Forming the inner products of both equations in Eq. 2.14 with the second basis ket $|\epsilon_2\rangle$ yields $A_{21} = \langle\epsilon_2|\,\widehat{A}\,|\epsilon_1\rangle$ and $A_{22} = \langle\epsilon_2|\,\widehat{A}\,|\epsilon_2\rangle$.

Combining these results gives the matrix representation of operator $\widehat{A}$ in a coordinate system with basis vectors represented by kets $|\epsilon_1\rangle$ and $|\epsilon_2\rangle$:

$$\bar{\bar{A}} = \begin{pmatrix}\langle\epsilon_1|\,\widehat{A}\,|\epsilon_1\rangle & \langle\epsilon_1|\,\widehat{A}\,|\epsilon_2\rangle \\ \langle\epsilon_2|\,\widehat{A}\,|\epsilon_1\rangle & \langle\epsilon_2|\,\widehat{A}\,|\epsilon_2\rangle\end{pmatrix}, \tag{2.15}$$

which can be written concisely as

$$A_{ij} = \langle\epsilon_i|\,\widehat{A}\,\big|\epsilon_j\big\rangle, \tag{2.16}$$

in which $|\epsilon_i\rangle$ and $\big|\epsilon_j\big\rangle$ represent a pair of orthonormal basis vectors.

Here's an example that shows how this might be useful. Consider the operator discussed in the previous section for which the matrix representation in the Cartesian coordinate system has elements given by $\bar{\bar{R}} = \begin{pmatrix}4 & -2 \\ -2 & 4\end{pmatrix}$. Imagine that you're interested in determining the elements of the matrix representing that operator in the two-dimensional orthonormal basis system with basis vectors $\hat{\epsilon}_1 = \frac{1}{\sqrt{2}}(\hat{\imath}+\hat{\jmath}) = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}$ and $\hat{\epsilon}_2 = \frac{1}{\sqrt{2}}(\hat{\imath}-\hat{\jmath}) = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}$.

Using Eq. 2.16, the elements of matrix $\bar{\bar{R}} = \begin{pmatrix} 4 & -2 \\ -2 & 4 \end{pmatrix}$ in the $(\hat{\epsilon}_1, \hat{\epsilon}_2)$ basis are found as

$$\begin{aligned} R_{11} = \langle \epsilon_1 | \widehat{R} | \epsilon_1 \rangle &= \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 4 & -2 \\ -2 & 4 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} (4)(\frac{1}{\sqrt{2}}) + (-2)(\frac{1}{\sqrt{2}}) \\ (-2)(\frac{1}{\sqrt{2}}) + (4)(\frac{1}{\sqrt{2}}) \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{2}{\sqrt{2}} \\ \frac{2}{\sqrt{2}} \end{pmatrix} \\ &= \frac{1}{\sqrt{2}} \frac{2}{\sqrt{2}} + \frac{1}{\sqrt{2}} \frac{2}{\sqrt{2}} = 2, \end{aligned}$$

$$\begin{aligned} R_{12} = \langle \epsilon_1 | \widehat{R} | \epsilon_2 \rangle &= \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 4 & -2 \\ -2 & 4 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} (4)(\frac{1}{\sqrt{2}}) + (-2)(-\frac{1}{\sqrt{2}}) \\ (-2)(\frac{1}{\sqrt{2}}) + (4)(-\frac{1}{\sqrt{2}}) \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{6}{\sqrt{2}} \\ -\frac{6}{\sqrt{2}} \end{pmatrix} \\ &= \frac{1}{\sqrt{2}} \frac{6}{\sqrt{2}} + \frac{1}{\sqrt{2}} \left(-\frac{6}{\sqrt{2}} \right) = 0, \end{aligned}$$

$$\begin{aligned} R_{21} = \langle \epsilon_2 | \widehat{R} | \epsilon_1 \rangle &= \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 4 & -2 \\ -2 & 4 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} (4)(\frac{1}{\sqrt{2}}) + (-2)(\frac{1}{\sqrt{2}}) \\ (-2)(\frac{1}{\sqrt{2}}) + (4)(\frac{1}{\sqrt{2}}) \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{2}{\sqrt{2}} \\ \frac{2}{\sqrt{2}} \end{pmatrix} \\ &= \frac{1}{\sqrt{2}} \frac{2}{\sqrt{2}} - \frac{1}{\sqrt{2}} \frac{2}{\sqrt{2}} = 0, \end{aligned}$$

and

$$\begin{aligned} R_{22} = \langle \epsilon_2 | \widehat{R} | \epsilon_2 \rangle &= \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 4 & -2 \\ -2 & 4 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} (4)(\frac{1}{\sqrt{2}}) + (-2)(-\frac{1}{\sqrt{2}}) \\ (-2)(\frac{1}{\sqrt{2}}) + (4)(-\frac{1}{\sqrt{2}}) \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{6}{\sqrt{2}} \\ -\frac{6}{\sqrt{2}} \end{pmatrix} \\ &= \frac{1}{\sqrt{2}} \frac{6}{\sqrt{2}} - \frac{1}{\sqrt{2}} \left(-\frac{6}{\sqrt{2}} \right) = 6. \end{aligned}$$

Thus

$$\bar{\bar{R}} = \begin{pmatrix} 2 & 0 \\ 0 & 6 \end{pmatrix} \qquad \text{in the } \hat{\epsilon}_1, \hat{\epsilon}_2 \text{ basis.}$$

If the values of the diagonal elements look familiar, it may be because they're the eigenvalues of matrix $\bar{\bar{R}}$, as found in the previous section. This is not a coincidence, because the basis vectors $\hat{\epsilon}_1 = \frac{1}{\sqrt{2}}(\hat{\imath} + \hat{\jmath})$ and $\hat{\epsilon}_2 = \frac{1}{\sqrt{2}}(\hat{\imath} - \hat{\jmath})$ are the (normalized) eigenvectors of this matrix. And when an operator matrix with nondegenerate eigenvalues (that is, no eigenvalue is shared by two or more eigenfunctions[4]) is expressed using its eigenfunctions as basis functions, the matrix is diagonal (that is, all off-diagonal elements are zero), and the diagonal elements are the eigenvalues of the matrix.

One additional bit of operator mathematics you're sure to encounter in your study of quantum mechanics is called "commutation." Two operators $\hat{A}$ and $\hat{B}$ are said to "commute" if the order of their application can be switched without changing the result. So operating on a ket $|\psi\rangle$ with operator $\hat{B}$ and then applying operator $\hat{A}$ to the result gives the same answer as first operating on $|\psi\rangle$ with operator $\hat{A}$ and then applying operator $\hat{B}$ to the result. This can be written as

$$\hat{A}(\hat{B}\,|\psi\rangle) = \hat{B}(\hat{A}\,|\psi\rangle) \quad \text{if } \widehat{A} \text{ and } \widehat{B} \text{ commute} \tag{2.17}$$

or

$$\hat{A}\hat{B}(|\psi\rangle) - \hat{B}\hat{A}(|\psi\rangle) = 0$$
$$\left(\hat{A}\hat{B} - \hat{B}\hat{A}\right)|\psi\rangle = 0.$$

The quantity in parenthesis $(\hat{A}\hat{B} - \hat{B}\hat{A})$ is called the commutator of operators $\hat{A}$ and $\hat{B}$ and is commonly written as

$$[\hat{A}, \hat{B}] = \hat{A}\hat{B} - \hat{B}\hat{A}. \tag{2.18}$$

So the bigger the change in the result caused by switching the order of operation, the bigger the commutator.

If you find it surprising that some pairs of operators don't commute, remember that operators can be represented by matrices, and matrix products are in general not commutative (that is, the order of multiplication matters).

To see an example of this, consider two matrices representing operators $\hat{A}$ and $\hat{B}$:

$$\bar{\bar{A}} = \begin{pmatrix} i & 0 & 1 \\ 0 & -i & 2 \\ 0 & -1 & 0 \end{pmatrix} \qquad \bar{\bar{B}} = \begin{pmatrix} 2 & i & 0 \\ 0 & 1 & -i \\ -1 & 0 & 0 \end{pmatrix}.$$

To determine whether these operators commute, compare the matrix product $\bar{\bar{A}}\bar{\bar{B}}$ to $\bar{\bar{B}}\bar{\bar{A}}$:

[4] You can read more about degenerate eigenvalues in the next section of this chapter.

$$\bar{\bar{A}}\bar{\bar{B}} = \begin{pmatrix} i & 0 & 1 \\ 0 & -i & 2 \\ 0 & -1 & 0 \end{pmatrix} \begin{pmatrix} 2 & i & 0 \\ 0 & 1 & -i \\ -1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 2i-1 & -1 & 0 \\ -2 & -i & -1 \\ 0 & -1 & i \end{pmatrix}$$

and

$$\bar{\bar{B}}\bar{\bar{A}} = \begin{pmatrix} 2 & i & 0 \\ 0 & 1 & -i \\ -1 & 0 & 0 \end{pmatrix} \begin{pmatrix} i & 0 & 1 \\ 0 & -i & 2 \\ 0 & -1 & 0 \end{pmatrix} = \begin{pmatrix} 2i & 1 & 2+2i \\ 0 & 0 & 2 \\ -i & 0 & -1 \end{pmatrix},$$

which means that matrices $\bar{\bar{A}}$ and $\bar{\bar{B}}$ (and their corresponding operators $\hat{A}$ and $\hat{B}$) do not commute. Subtracting gives the commutator $[\bar{\bar{A}}, \bar{\bar{B}}]$:

$$[\bar{\bar{A}}, \bar{\bar{B}}] = \bar{\bar{A}}\bar{\bar{B}} - \bar{\bar{B}}\bar{\bar{A}} = \begin{pmatrix} -1 & -2 & -2-2i \\ -2 & -i & -3 \\ i & -1 & 1+i \end{pmatrix}.$$

Main Ideas of This Section

The elements of the matrix representation of an operator in a specified basis may be determined by sandwiching the operator between pairs of basis vectors. Two operators for which changing the order of operation does not change the result are said to commute.

Relevance to Quantum Mechanics

In Section 2.4, the elements of the matrix representing the important quantum operator called the "projection operator" will be found by sandwiching that operator between pairs of basis vectors. In Section 2.5, you'll see that the expression $\langle\psi|\,\widehat{O}\,|\psi\rangle$ can be used to determine the expectation value of measurements of the quantum observable corresponding to operator $\widehat{O}$ for a system in state $|\psi\rangle$.

Every quantum observable has an associated operator, and if two operators commute, the measurements associated with those two operators may be done in either order with the same result. That means that those two observables may be simultaneously known with precision limited only by the experimental configuration and instrumentation, whereas the Heisenberg Uncertainty Principle limits the precision with which two observables whose operators do not commute may be simultaneously known.

2.3 Hermitian Operators

An important characteristic of quantum operators may be understood by considering both sides of Eq. 2.11:

$$\langle\phi|\,\widehat{O}\,|\psi\rangle = \langle\phi|\,\lambda\,|\psi\rangle\,, \tag{2.11}$$

in which $|\phi\rangle$ and $|\psi\rangle$ represent quantum wavefunctions.

The right side of Eq. 2.11 is easy to deal with, since the constant λ is outside both bra $\langle\phi|$ and ket $|\psi\rangle$. That constant may be moved either to the right of the ket or to the left of the bra, so the expression $\langle\phi|\,\lambda\,|\psi\rangle$ can be written as

$$\langle\phi|\,\lambda\,|\psi\rangle = \langle\phi|\psi\rangle\,\lambda = \lambda\,\langle\phi|\psi\rangle\,, \tag{2.19}$$

which you'll see again later in this section. But it's the left side of Eq. 2.11 that contains some interesting and useful concepts.

As mentioned in the previous section, the operator $\widehat{O}$ in Eq. 2.11 can operate either to the right on ket $|\psi\rangle$ or to the left on bra $\langle\phi|$ (the dual of ket $|\phi\rangle$), so you can think of this expression in either of two ways. One is

$$\langle\phi| \longrightarrow \quad \widehat{O}\,|\psi\rangle\,,$$

in which bra $\langle\phi|$ is headed toward an encounter with the ket that results from the operation of $\widehat{O}$ on ket $|\psi\rangle$. That encounter takes the form of an inner product.

Alternatively, you can view Eq. 2.11 like this:

$$\langle\phi|\,\widehat{O} \longrightarrow \quad |\psi\rangle$$

in which bra $\langle\phi|$ is operated upon by $\widehat{O}$ and the result (another bra, remember) is destined to run into ket $|\psi\rangle$.

Both of these perspectives are valid, and you'll get the same result no matter which way you apply the operator, as long as you apply that operator correctly. There's a bit of subtlety involved in the second approach, and that subtlety involves adjoints and Hermitian operators, which are important topics in their own right.

The first approach (operating $\widehat{O}$ on $|\psi\rangle$) is straightforward. If you'd like, you can move operator $\widehat{O}$ right inside the ket brackets with the label ψ, making a new ket:

$$\widehat{O}\,|\psi\rangle = |\widehat{O}\psi\rangle\,. \tag{2.20}$$

It may seem strange to see an operator inside a ket, since up to this point we've considered operators as operating *on* kets rather than within them. But remember that the symbol inside the ket such as ψ or $\widehat{O}\psi$ is just a

label – specifically, it's the name of the vector represented by the ket. So when you move an operator into a ket to make a new ket (such as $|\widehat{O}\psi\rangle$), what you're really doing is changing the vector to which the ket refers, from vector $\vec{\psi}$ to the vector produced by operating $\widehat{O}$ on $\vec{\psi}$. If you give that new vector the name $\overrightarrow{\widehat{O}\psi}$, then the associated ket is $|\widehat{O}\psi\rangle$. It's that new ket that forms an inner product with $|\phi\rangle$ in the expression $\langle\phi|\,\widehat{O}\,|\psi\rangle$.

Going to the left with the operator in an expression such as $\langle\phi|\,\widehat{O}\,|\psi\rangle$ can be done in two ways, one of which involves moving the operator $\widehat{O}$ inside the bra $\langle\phi|$. But you can't move an operator inside a bra without changing that operator. That change is called taking the "adjoint" of the operator,[5] written as $\widehat{O}^\dagger$. So the process of moving operator $\widehat{O}$ from outside to inside a bra looks like this:

$$\langle\psi|\,\widehat{O} = \langle\widehat{O}^\dagger\psi|\,. \tag{2.21}$$

When you consider the expression $\langle\widehat{O}^\dagger\psi|$, remember that the label inside a bra (such as $\widehat{O}^\dagger\psi$) refers to a vector – in this case, the vector that is formed by allowing operator $\widehat{O}^\dagger$ to operate on vector $\vec{\psi}$. So the bra $\langle\widehat{O}^\dagger\psi|$ is the dual of ket $|\widehat{O}^\dagger\psi\rangle$.

Finding the adjoint of an operator in matrix form is straightforward. Just take the complex conjugate of each element of the matrix, and then form the transpose of the matrix – that is, interchange the rows and columns of the matrix. So the first row becomes the first column, the second row becomes the second column, and so forth. If operator $\widehat{O}$ has matrix representation

$$\widehat{O} = \begin{pmatrix} O_{11} & O_{12} & O_{13} \\ O_{21} & O_{22} & O_{23} \\ O_{31} & O_{32} & O_{33} \end{pmatrix}, \tag{2.22}$$

then its adjoint $\widehat{O}^\dagger$ is

$$\widehat{O}^\dagger = \begin{pmatrix} O_{11}^* & O_{21}^* & O_{31}^* \\ O_{12}^* & O_{22}^* & O_{32}^* \\ O_{13}^* & O_{23}^* & O_{33}^* \end{pmatrix}. \tag{2.23}$$

If you think about applying this conjugate-transpose process to a column vector, you'll see that the Hermitian adjoint of a ket is the associated bra:

$$|A\rangle = \begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix}$$

$$|A\rangle^\dagger = \begin{pmatrix} A_1^* & A_2^* & A_3^* \end{pmatrix} = \langle A|\,.$$

[5] Also called the "transpose conjugate" or "Hermitian conjugate" of the operator.

It's useful to know how $\widehat{O}$ and its adjoint $\widehat{O}^\dagger$ differ in form, but you should also understand how they differ in function. Here's the answer: if $\widehat{O}$ transforms ket $|\psi\rangle$ into ket $|\psi'\rangle$, then $\widehat{O}^\dagger$ transforms bra $\langle\psi|$ into bra $\langle\psi'|$. In equations this is

$$\begin{aligned} \widehat{O}\,|\psi\rangle &= |\psi'\rangle \\ \langle\psi|\,\widehat{O}^\dagger &= \langle\psi'| \,, \end{aligned} \tag{2.24}$$

in which bra $\langle\psi|$ is the dual of ket $|\psi\rangle$ and bra $\langle\psi'|$ is the dual of ket $|\psi'\rangle$. Be sure to note that in Eqs. 2.24 the operators $\widehat{O}$ and $\widehat{O}^\dagger$ are outside $|\psi\rangle$ and $\langle\psi|$.

You should also be aware that it's perfectly acceptable to evaluate an expression such as $\langle\psi|\,\widehat{O}$ without moving the operator inside the bra. Since a bra can be represented by a row vector, a bra standing on the left of an operator can be written as a row vector standing on the left of a matrix. That means you can multiply them together as long as the number of elements in the row vector matches the number of rows in the matrix. So if $|\psi\rangle$, $\langle\psi|$, and $\widehat{O}$ are given by

$$|\psi\rangle = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} \qquad \langle\psi| = \begin{pmatrix} \psi_1^* & \psi_2^* \end{pmatrix} \qquad \widehat{O} = \begin{pmatrix} O_{11} & O_{12} \\ O_{21} & O_{22} \end{pmatrix},$$

then

$$\begin{aligned} \langle\psi|\,\widehat{O} &= \begin{pmatrix} \psi_1^* & \psi_2^* \end{pmatrix} \begin{pmatrix} O_{11} & O_{12} \\ O_{21} & O_{22} \end{pmatrix} \\ &= \begin{pmatrix} \psi_1^* O_{11} + \psi_2^* O_{21} & \psi_1^* O_{12} + \psi_2^* O_{22} \end{pmatrix}, \end{aligned} \tag{2.25}$$

which is the same result as $\langle\widehat{O}^\dagger\psi|$:

$$\begin{aligned} \widehat{O}^\dagger &= \begin{pmatrix} O_{11}^* & O_{21}^* \\ O_{12}^* & O_{22}^* \end{pmatrix} \\ \langle\widehat{O}^\dagger\psi| = |\widehat{O}^\dagger\psi\rangle^\dagger &= \left(\widehat{O}^\dagger\,|\psi\rangle\right)^\dagger = \left[\begin{pmatrix} O_{11}^* & O_{21}^* \\ O_{12}^* & O_{22}^* \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}\right]^\dagger \\ &= \begin{pmatrix} \psi_1 O_{11}^* + \psi_2 O_{21}^* \\ \psi_1 O_{12}^* + \psi_2 O_{22}^* \end{pmatrix}^\dagger \\ &= \begin{pmatrix} \psi_1^* O_{11} + \psi_2^* O_{21} & \psi_1^* O_{12} + \psi_2^* O_{22} \end{pmatrix}, \end{aligned} \tag{2.26}$$

in agreement with Eq. 2.25.

So when you're confronted with a bra standing to the left of an operator (outside the bra), you can either multiply the row vector representing the bra by the matrix representing the operator, or you can move the operator into the bra, taking the operator's Hermitian conjugate in the process.

With an understanding of how to deal with operators outside and inside bras and kets, you should be able to see the equivalence of the following expressions:

$$\langle\phi|\,\widehat{O}\,|\psi\rangle = \langle\phi|\widehat{O}\psi\rangle = \langle\widehat{O}^{\dagger}\phi|\psi\rangle\,. \tag{2.27}$$

The reason for making the effort to get to Eq. 2.27 is to help you understand an extremely important characteristic of certain operators. Those operators are called "Hermitian," and their defining characteristic is this: Hermitian operators equal their own adjoints. So if $\widehat{O}$ is a Hermitian operator, then

$$\widehat{O} = \widehat{O}^{\dagger} \qquad \text{(Hermitian } \widehat{O}\text{)}. \tag{2.28}$$

It's easy to determine whether an operator is Hermitian by looking at the operator's matrix representation. Comparing Eqs. 2.22 and 2.23, you can see that for a matrix to equal its own adjoint, the diagonal elements must all be real (since only a purely real number equals its complex conjugate), and every off-diagonal element must equal the complex conjugate of the corresponding element on the other side of the diagonal (so O_{21} must equal O_{12}^{*}, O_{31} must equal O_{13}^{*}, O_{23} must equal O_{32}^{*}, and so forth).

Why are Hermitian operators of special interest? To see that, look again at the second equality in Eq. 2.27. If operator $\widehat{O}$ equals its adjoint $\widehat{O}^{\dagger}$, then

$$\langle\phi|\,\widehat{O}\,|\psi\rangle = \langle\phi|\widehat{O}\psi\rangle = \langle\widehat{O}^{\dagger}\phi|\psi\rangle = \langle\widehat{O}\phi|\psi\rangle\,, \tag{2.29}$$

which means that a Hermitian operator may be applied to *either* member of an inner product with the same result.

For complex continuous functions such as $f(x)$ and $g(x)$, the equivalent to Eq. 2.29 is

$$\begin{aligned}\int_{\infty}^{\infty} f^{*}(x)\left[\widehat{O}g(x)\right]dx &= \int_{\infty}^{\infty}\left[\widehat{O}^{\dagger}f^{*}(x)\right]g(x)dx \\ &= \int_{\infty}^{\infty}\left[\widehat{O}f^{*}(x)\right]g(x)dx.\end{aligned} \tag{2.30}$$

The ability to move a Hermitian operator to either side of an inner product may seem like a minor computational benefit, but it has major ramifications. To appreciate those ramifications, consider what happens when a Hermitian operator is sandwiched between a ket such as $|\psi\rangle$ and its corresponding bra $\langle\psi|$. That makes Eq. 2.29

$$\langle\psi|\,\widehat{O}\,|\psi\rangle = \langle\psi|\widehat{O}\psi\rangle = \langle\widehat{O}\psi|\psi\rangle\,. \tag{2.31}$$

Now consider what this equation means if $|\psi\rangle$ is an eigenket of $\widehat{O}$ with eigenvalue λ. In that case, $\left|\widehat{O}\psi\right\rangle = |\lambda\psi\rangle$ and $\left\langle\widehat{O}\psi\right| = \langle\lambda\psi|$, so

$$\langle\psi|\lambda\psi\rangle = \langle\lambda\psi|\psi\rangle\,. \tag{2.32}$$

To learn something from this equation, you need to understand the rules for pulling a constant from inside to outside (or outside to inside) a ket or bra. For kets, you can move a constant, even if that constant is complex, from inside to outside (or outside to inside) a ket without changing the constant. So

$$c\,|A\rangle = |cA\rangle\,. \tag{2.33}$$

You can see why this is true by writing the ket as a column vector:

$$c\,|A\rangle = c\begin{pmatrix}A_\mathrm{x}\\A_\mathrm{y}\\A_\mathrm{z}\end{pmatrix} = \begin{pmatrix}cA_\mathrm{x}\\cA_\mathrm{y}\\cA_\mathrm{z}\end{pmatrix} = |cA\rangle\,.$$

But if you want to move a constant from inside to outside (or outside to inside) a bra, it's necessary to take the complex conjugate of that constant:

$$c\,\langle A| = \left\langle c^*A\right|, \tag{2.34}$$

because in this case

$$\begin{aligned} c\,\langle A| = c\begin{pmatrix}A_\mathrm{x}^* & A_\mathrm{y}^* & A_\mathrm{z}^*\end{pmatrix} &= \begin{pmatrix}cA_\mathrm{x}^* & cA_\mathrm{y}^* & cA_\mathrm{z}^*\end{pmatrix} \\ &= \begin{pmatrix}(c^*A_\mathrm{x})^* & (c^*A_\mathrm{y})^* & (c^*A_\mathrm{z})^*\end{pmatrix} = \left\langle c^*A\right|. \end{aligned}$$

If you don't see why that last equality is true, remember that for the ket

$$\left|c^*A\right\rangle = \begin{pmatrix}c^*A_\mathrm{x}\\c^*A_\mathrm{y}\\c^*A_\mathrm{z}\end{pmatrix},$$

the corresponding bra is $\langle c^*A| = \begin{pmatrix}(c^*A_\mathrm{x})^* & (c^*A_\mathrm{y})^* & (c^*A_\mathrm{z})^*\end{pmatrix}$. This matches the expression for $c\,\langle A|$, so $c\,\langle A| = \langle c^*A|$.

The result of all this is that a constant can be moved in or out of a ket without change, but moving a constant in or out of a bra requires you to take the complex conjugate of the constant. So pulling the constant λ out of the ket $|\lambda\psi\rangle$ on the left side of Eq. 2.32 and out of the bra $\langle\lambda\psi|$ on the right side of that equation gives

$$\langle\psi|\,\lambda\,|\psi\rangle = \lambda^*\,\langle\psi|\psi\rangle\,. \tag{2.35}$$

At the start of this section, you saw that a constant sandwiched between a bra and a ket (but not inside either one) can be moved either to the left of the bra

or to the right of the ket without change. Pulling the constant λ from between bra $\langle\psi|$ and ket $|\psi\rangle$ on the left side of Eq. 2.35 gives

$$\lambda \langle\psi|\psi\rangle = \lambda^* \langle\psi|\psi\rangle . \tag{2.36}$$

This can be true only if $\lambda = \lambda^*$, which means that the eigenvalue λ must be real. So Hermitian operators must have real eigenvalues.

Another useful result can be obtained by considering an expression in which a Hermitian operator is sandwiched between two different functions, as in Eq. 2.29:

$$\langle\phi|\,\widehat{O}\,|\psi\rangle = \langle\phi|\widehat{O}\psi\rangle = \langle\widehat{O}^\dagger\phi|\psi\rangle = \langle\widehat{O}\phi|\psi\rangle . \tag{2.29}$$

Consider the case in which ϕ is an eigenfunction of Hermitian operator $\widehat{O}$ with eigenvalue λ_ϕ and ψ is also an eigenfunction of $\widehat{O}$ with (different) eigenvalue λ_ψ. Eq. 2.29 is then

$$\langle\phi|\,\widehat{O}\,|\psi\rangle = \langle\phi|\lambda_\psi\,\psi\rangle = \langle\lambda_\phi\phi|\psi\rangle ,$$

and pulling out the constants λ_ψ and λ_ϕ gives

$$\lambda_\psi \langle\phi|\psi\rangle = \lambda_\phi^* \langle\phi|\psi\rangle .$$

But the eigenvalues of Hermitian operators must be real, so $\lambda_\phi^* = \lambda_\phi$, and

$$\begin{aligned} \lambda_\psi \langle\phi|\psi\rangle &= \lambda_\phi \langle\phi|\psi\rangle \\ (\lambda_\psi - \lambda_\phi) \langle\phi|\psi\rangle &= 0. \end{aligned}$$

This means that either $(\lambda_\psi - \lambda_\phi)$ or $\langle\phi|\psi\rangle$ (or both) must be zero. But we specified that the eigenfunctions ϕ and ψ have different eigenvalues, so $(\lambda_\psi - \lambda_\phi)$ cannot be zero, and the only possibility is that $\langle\phi|\psi\rangle = 0$. Since the inner product between two functions can be zero only when the functions are orthogonal, this means that the eigenfunctions of a Hermitian operator with different eigenvalues must be orthogonal.

And what if two or more eigenfunctions share an eigenvalue? That's called the "**degenerate**" case, and the eigenfunctions with the same eigenvalue will not, in general, be orthogonal. But in such cases it is always possible to use a weighted combination of the non-orthogonal eigenfunctions to produce an orthogonal set of eigenfunctions for the degenerate eigenvalue. So in the nondegenerate case (in which no eigenfunctions share an eigenvalue), only one set of eigenfunctions exist, and those eigenfunctions are guaranteed to be orthogonal. But in the degenerate case, there are an infinite number of non-orthogonal eigenfunctions, from which you can always construct an orthogonal set.[6]

[6] The **Gram–Schmidt procedure** for constructing a set of orthogonal vectors is explained on the book's website.

There's one more useful characteristic of the eigenfuctions of a **Hermitian operator**: they form a complete set. That means that any function in the abstract vector space containing the eigenfunctions of a Hermitian operator may be made up of a linear combination of those eigenfunctions.

Main Ideas of This Section
Hermitian operators may be applied to either member of an inner product and the result will be the same. Hermitian operators have real eigenvalues, and the nondegenerate eigenfunctions of a Hermitian operator are orthogonal and form a complete set.

Relevance to Quantum Mechanics
The discussion of the solutions to the Schrödinger equation in Chapter 4 will show that every quantum observable (such as position, momentum, and energy) is associated with an operator, and the possible results of any measurement are given by the eigenvalues of that operator. Since the results of measurements must be real, operators associated with observables must be Hermitian. The eigenfunctions of Hermitian operators are (or can be combined to be) orthogonal, and the orthogonality of those eigenfunctions has a profound impact on our ability to construct solutions to the Schrödinger equation and to use those solutions to determine the probability of various measurement outcomes.

2.4 Projection Operators

A very useful Hermitian operator you're likely to encounter in most books on quantum mechanics is the "projection operator." To understand what's being projected, consider the ket representing three-dimensional vector $\vec{A}$. Expanding that ket using the basis kets representing orthonormal vectors $\hat{\epsilon}_1$, $\hat{\epsilon}_2$, and $\hat{\epsilon}_3$ looks like this:

$$|A\rangle = A_1 |\epsilon_1\rangle + A_2 |\epsilon_2\rangle + A_3 |\epsilon_3\rangle . \tag{2.37}$$

Or, using Eq. 1.32 for the components A_1, A_2, and A_3, like this:

$$|A\rangle = \langle\epsilon_1|A\rangle |\epsilon_1\rangle + \langle\epsilon_2|A\rangle |\epsilon_2\rangle + \langle\epsilon_3|A\rangle |\epsilon_3\rangle \tag{2.38}$$

since $\langle\epsilon_i|\epsilon_i\rangle = 1$ for orthonormal basis vectors.

And since the inner products $\langle\epsilon_1|A\rangle$, $\langle\epsilon_2|A\rangle$, and $\langle\epsilon_3|A\rangle$ are scalars (as they must be, since they represent A_1, A_2, and A_3), you can move them to the other side of the basis kets $|\epsilon_1\rangle$, $|\epsilon_2\rangle$, and $|\epsilon_3\rangle$, and the expansion of the ket representing $\vec{A}$ becomes

$$|A\rangle = |\epsilon_1\rangle\,\langle\epsilon_1|A\rangle + |\epsilon_2\rangle\,\langle\epsilon_2|A\rangle + |\epsilon_3\rangle\,\langle\epsilon_3|A\rangle\,. \tag{2.39}$$

This equation came about with the terms grouped as

$$|A\rangle = |\epsilon_1\rangle\,\underbrace{\langle\epsilon_1|A\rangle}_{A_1} + |\epsilon_2\rangle\,\underbrace{\langle\epsilon_2|A\rangle}_{A_2} + |\epsilon_3\rangle\,\underbrace{\langle\epsilon_3|A\rangle}_{A_3},$$

but consider the alternative grouping

$$|A\rangle = \underbrace{|\epsilon_1\rangle\,\langle\epsilon_1|}_{\widehat{P}_1}\;|A\rangle + \underbrace{|\epsilon_2\rangle\,\langle\epsilon_2|}_{\widehat{P}_2}\;|A\rangle + \underbrace{|\epsilon_3\rangle\,\langle\epsilon_3|}_{\widehat{P}_3}\;|A\rangle\,. \tag{2.40}$$

As you can see from the labels underneath the curly braces, the terms $|\epsilon_1\rangle\,\langle\epsilon_1|$, $|\epsilon_2\rangle\,\langle\epsilon_2|$, and $|\epsilon_3\rangle\,\langle\epsilon_3|$ are the operators $\widehat{P}_1$, $\widehat{P}_2$, and $\widehat{P}_3$.

The general expression for a projection operator is

$$\widehat{P}_i = |\epsilon_i\rangle\,\langle\epsilon_i|\,, \tag{2.41}$$

in which $\hat{\epsilon}_i$ is any normalized vector. This expression, with a ket standing to the left of a bra, may look a bit strange at first, but most operators look strange until you feed them something on which to operate. Feeding operator $\widehat{P}_1$ the ket representing vector $\vec{A}$ helps you see what's happening:

$$\widehat{P}_1\,|A\rangle = |\epsilon_1\rangle\,\langle\epsilon_1|A\rangle = A_1\,|\epsilon_1\rangle \tag{2.42}$$

So applying the projection operator to $|A\rangle$ produces the new ket $A_1\,|\epsilon_1\rangle$. The magnitude of that new ket is the (scalar) projection of the ket that you feed into the operator (in this case, $|A\rangle$) onto the direction of the ket you use to define the operator (in this case, $|\epsilon_1\rangle$). But here's an important step: that magnitude is then multiplied by the ket you use to define the operator. So the result of applying the projection operator to a ket is not just the (scalar) component (such as A_1) of that ket along the direction of $\hat{\epsilon}_1$, it's a new ket in that direction. Put in terms of a vector in the Cartesian coordinate system, the $\widehat{P}_1$ projection operator doesn't just give you the scalar A_x, it gives you the vector $A_x\hat{\imath}$.

In defining a projection operator, it's necessary to use a ket representing a normalized vector (such as $\hat{\epsilon}_1$) within the operator; you can think of that vector as the "projector vector." If the projector vector doesn't have unit length, then its length contributes to the result of the inner product as well as the result

of the multiplication by the projector vector. To remove those contributions requires dividing by the square of the norm of the (non-normalized) projector vector.[7]

For completeness, the results of applying the three projection operators $\widehat{P}_1$, $\widehat{P}_2$, and $\widehat{P}_3$ to ket $|A\rangle$ are

$$\begin{aligned}\widehat{P}_1 |A\rangle &= |\epsilon_1\rangle \langle\epsilon_1|A\rangle = A_1 |\epsilon_1\rangle \\ \widehat{P}_2 |A\rangle &= |\epsilon_2\rangle \langle\epsilon_2|A\rangle = A_2 |\epsilon_2\rangle \\ \widehat{P}_3 |A\rangle &= |\epsilon_3\rangle \langle\epsilon_3|A\rangle = A_3 |\epsilon_3\rangle .\end{aligned} \tag{2.43}$$

If you sum the results of applying the projection operators for all of the basis kets in a three-dimensional space, the result is

$$\widehat{P}_1 |A\rangle + \widehat{P}_2 |A\rangle + \widehat{P}_3 |A\rangle = A_1 |\epsilon_1\rangle + A_2 |\epsilon_2\rangle + A_3 |\epsilon_3\rangle = |A\rangle$$

or

$$\left(\widehat{P}_1 + \widehat{P}_2 + \widehat{P}_3\right) |A\rangle = |A\rangle .$$

Writing this for the general case in an N-dimensional space:

$$\sum_{n=1}^{N} \widehat{P}_n |A\rangle = |A\rangle . \tag{2.44}$$

This means that the sum of the projection operators using all of the basis vectors equals the "identity operator" $\widehat{I}$. The identity operator is the Hermitian operator that produces a ket that is equal to the ket that is fed into the operator:

$$\widehat{I} |A\rangle = |A\rangle . \tag{2.45}$$

This works for any ket, not only $|A\rangle$, just as multiplying any number by the number "1" produces the same number. The matrix representation ($\bar{\bar{I}}$) of the identity operator in three dimensions is

$$\bar{\bar{I}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} . \tag{2.46}$$

The relation

$$\sum_{n=1}^{N} \widehat{P}_n = \sum_{n=1}^{N} |\epsilon_n\rangle \langle\epsilon_n| = \widehat{I} \tag{2.47}$$

[7] That's why you'll see projection operators defined as $\widehat{P}_i = \frac{|\epsilon_i\rangle\langle\epsilon_i|}{|\epsilon_i|^2}$ in some texts.

is called the "completeness" or **"closure" relation**, since it holds true when applied to any ket in an N-dimensional space. That means that any ket in that space can be represented as the sum of N basis kets weighted by N components. In other words, the basis vectors $\vec{\epsilon}_n$ represented by the kets $|\epsilon_n\rangle$, and their dual bras $\langle\epsilon_n|$ in Eq. 2.47 form a complete set.

Like all operators, the projection operator in an N-dimensional space may be represented by an N×N matrix. You can find the elements of that matrix using Eq. 2.16:

$$A_{ij} = \langle\epsilon_i| \widehat{A} \left|\epsilon_j\right\rangle . \tag{2.16}$$

As explained in Section 2.2, before you can find the elements of the matrix representation of an operator, it's necessary to decide which basis system you'd like to use (just as you need to decide on a basis system before finding the components of a vector).

One option is to use the basis system consisting of the eigenkets of the operator. As you may recall, in that basis the matrix representing an operator is diagonal, and each of the diagonal elements is an eigenvalue of the matrix.

Finding the eigenkets and eigenvalues of the projection operator is straightforward. For projection operator $\widehat{P}_1$, for example, the eigenket equation is

$$\widehat{P}_1 |A\rangle = \lambda_1 |A\rangle , \tag{2.48}$$

in which ket $|A\rangle$ is an eigenket of $\widehat{P}_1$ with eigenvalue λ_1. Inserting $|\epsilon_1\rangle \langle\epsilon_1|$ for $\widehat{P}_1$ gives

$$|\epsilon_1\rangle \langle\epsilon_1|A\rangle = \lambda_1 |A\rangle .$$

To see if the basis ket $|\epsilon_1\rangle$ is itself an eigenket of $\widehat{P}_1$, let $|A\rangle = |\epsilon_1\rangle$:

$$|\epsilon_1\rangle \langle\epsilon_1|\epsilon_1\rangle = \lambda_1 |\epsilon_1\rangle .$$

But $|\epsilon_1\rangle$, $|\epsilon_2\rangle$, and $|\epsilon_3\rangle$ form an orthonormal set, so $\langle\epsilon_1|\epsilon_1\rangle = 1$, which means

$$\begin{aligned} |\epsilon_1\rangle (1) &= \lambda_1 |\epsilon_1\rangle \\ 1 &= \lambda_1 . \end{aligned}$$

Hence $|\epsilon_1\rangle$ is indeed an eigenket of $\widehat{P}_1$, and the eigenvalue for this eigenket is one.

Having succeeded with $|\epsilon_1\rangle$ as an eigenket of $\widehat{P}_1$, let's try $|\epsilon_2\rangle$:

$$|\epsilon_1\rangle \langle\epsilon_1|\epsilon_2\rangle = \lambda_2 |\epsilon_2\rangle .$$

But $\langle \epsilon_1 | \epsilon_2 \rangle = 0$, so

$$|\epsilon_1\rangle (0) = \lambda_2 |\epsilon_1\rangle$$
$$0 = \lambda_2,$$

which means that $|\epsilon_2\rangle$ is also an eigenket of $\widehat{P}_1$; in this case the eigenvalue is zero. Similar analysis applied to $|\epsilon_3\rangle$ reveals that $|\epsilon_3\rangle$ is also an eigenket of $\widehat{P}_1$; its eigenvalue is also zero.

So the eigenkets of operator $\widehat{P}_1$ are $|\epsilon_1\rangle$, $|\epsilon_2\rangle$, and $|\epsilon_3\rangle$, with eigenvalues of 1, 0, and 0, respectively. With these eigenkets in hand, the matrix elements $(P_1)_{ij}$ can be found by inserting $\widehat{P}_1$ into Eq. 2.16:

$$(P_1)_{ij} = \langle \epsilon_i | \widehat{P}_1 \left| \epsilon_j \right\rangle . \tag{2.49}$$

Setting $i = 1$ and $j = 1$ and using $\widehat{P}_1 = |\epsilon_1\rangle \langle \epsilon_1|$ gives $(P_1)_{11}$:

$$(P_1)_{11} = \langle \epsilon_1 | \widehat{P}_1 | \epsilon_1 \rangle = \langle \epsilon_1 | \epsilon_1 \rangle \langle \epsilon_1 | \epsilon_1 \rangle = (1)(1) = 1.$$

Likewise

$$(P_1)_{12} = \langle \epsilon_1 | \widehat{P}_1 | \epsilon_2 \rangle = \langle \epsilon_1 | \epsilon_1 \rangle \langle \epsilon_1 | \epsilon_2 \rangle = (1)(0) = 0$$
$$(P_1)_{21} = \langle \epsilon_2 | \widehat{P}_1 | \epsilon_1 \rangle = \langle \epsilon_2 | \epsilon_1 \rangle \langle \epsilon_1 | \epsilon_1 \rangle = (0)(1) = 0$$
$$(P_1)_{13} = \langle \epsilon_1 | \widehat{P}_1 | \epsilon_3 \rangle = \langle \epsilon_1 | \epsilon_1 \rangle \langle \epsilon_1 | \epsilon_3 \rangle = (1)(0) = 0$$
$$(P_1)_{31} = \langle \epsilon_3 | \widehat{P}_1 | \epsilon_1 \rangle = \langle \epsilon_3 | \epsilon_1 \rangle \langle \epsilon_1 | \epsilon_1 \rangle = (0)(1) = 0$$
$$(P_1)_{23} = \langle \epsilon_2 | \widehat{P}_1 | \epsilon_3 \rangle = \langle \epsilon_2 | \epsilon_1 \rangle \langle \epsilon_1 | \epsilon_3 \rangle = (0)(0) = 0$$
$$(P_1)_{32} = \langle \epsilon_3 | \widehat{P}_1 | \epsilon_2 \rangle = \langle \epsilon_3 | \epsilon_1 \rangle \langle \epsilon_1 | \epsilon_2 \rangle = (0)(0) = 0.$$

Thus the matrix representing operator $\widehat{P}_1$ in the basis of its eigenkets $|\epsilon_1\rangle$, $|\epsilon_2\rangle$, and $|\epsilon_3\rangle$ is

$$\bar{\bar{P}}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{2.50}$$

As expected, in this basis the $\widehat{P}_1$ matrix is diagonal with diagonal elements equal to the eigenvalues 1, 0, and 0.

A similar analysis for the projection operator $\widehat{P}_2 = |\epsilon_2\rangle \langle \epsilon_2|$ shows that it has the same eigenkets ($|\epsilon_1\rangle$, $|\epsilon_2\rangle$, and $|\epsilon_3\rangle$) as $\widehat{P}_1$, with eigenvalues of 0, 1, and 0. Its matrix representation is therefore

$$\bar{\bar{P}}_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{2.51}$$

And projection operator $\widehat{P}_3 = |\epsilon_3\rangle\langle\epsilon_3|$ has the same eigenkets with eigenvalues of 0, 0, and 1; its matrix representation is

$$\bar{\bar{P}}_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{2.52}$$

According to the completeness relation (Eq. 2.47), the matrix representations of the projection operators $\widehat{P}_1$, $\widehat{P}_2$, and $\widehat{P}_3$ should add up to the matrix of the identity operator, and they do:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \bar{\bar{I}}. \tag{2.53}$$

An alternative method of finding the matrix elements of the projection operator $\widehat{P}_1$ is to use the outer product rule for matrix multiplication. That rule says that the outer product of a column vector $\vec{A}$ and a row vector $\vec{B}$ is

$$\begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix} \begin{pmatrix} B_1 & B_2 & B_3 \end{pmatrix} = \begin{pmatrix} A_1B_1 & A_1B_2 & A_1B_3 \\ A_2B_1 & A_2B_2 & A_2B_3 \\ A_3B_1 & A_3B_2 & A_3B_3 \end{pmatrix}. \tag{2.54}$$

Recall from Section 1.2 that you can expand basis vectors in their own "standard" basis system, in which case each vector will have a single nonzero component (and that component will equal one if the basis is orthonormal). So expanding the kets $|\epsilon_1\rangle$, $|\epsilon_2\rangle$, and $|\epsilon_3\rangle$ in their own basis makes them and their corresponding bras look like this:

$$|\epsilon_1\rangle = 1\,|\epsilon_1\rangle + 0\,|\epsilon_2\rangle + 0\,|\epsilon_3\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \qquad \langle\epsilon_1| = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}$$

$$|\epsilon_2\rangle = 0\,|\epsilon_1\rangle + 1\,|\epsilon_2\rangle + 0\,|\epsilon_3\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \qquad \langle\epsilon_2| = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix}$$

$$|\epsilon_3\rangle = 0\,|\epsilon_1\rangle + 0\,|\epsilon_2\rangle + 1\,|\epsilon_3\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \qquad \langle\epsilon_3| = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}.$$

With the outer product definition of Eq. 2.54 and these expressions for the basis kets and bras, the elements of the projection operators $\widehat{P}_1$, $\widehat{P}_2$, and $\widehat{P}_3$ can be found:

$$\widehat{P}_1 = |\epsilon_1\rangle\langle\epsilon_1| = \begin{pmatrix}1\\0\\0\end{pmatrix}\begin{pmatrix}1 & 0 & 0\end{pmatrix}$$
$$= \begin{pmatrix}(1)(1) & (1)(0) & (1)(0)\\(0)(1) & (0)(0) & (0)(0)\\(0)(1) & (0)(0) & (0)(0)\end{pmatrix} = \begin{pmatrix}1 & 0 & 0\\0 & 0 & 0\\0 & 0 & 0\end{pmatrix}$$

$$\widehat{P}_2 = |\epsilon_2\rangle\langle\epsilon_2| = \begin{pmatrix}0\\1\\0\end{pmatrix}\begin{pmatrix}0 & 1 & 0\end{pmatrix}$$
$$= \begin{pmatrix}(0)(0) & (0)(1) & (0)(0)\\(1)(0) & (1)(1) & (1)(0)\\(0)(0) & (0)(1) & (0)(0)\end{pmatrix} = \begin{pmatrix}0 & 0 & 0\\0 & 1 & 0\\0 & 0 & 0\end{pmatrix}$$

$$\widehat{P}_3 = |\epsilon_3\rangle\langle\epsilon_3| = \begin{pmatrix}0\\0\\1\end{pmatrix}\begin{pmatrix}0 & 0 & 1\end{pmatrix}$$
$$= \begin{pmatrix}(0)(0) & (0)(0) & (0)(1)\\(0)(0) & (0)(0) & (0)(1)\\(1)(0) & (1)(0) & (1)(1)\end{pmatrix} = \begin{pmatrix}0 & 0 & 0\\0 & 0 & 0\\0 & 0 & 1\end{pmatrix}.$$

You can see how to use the matrix outer product to find the elements of projector operators in other basis systems in the chapter-end problems and online solutions.

Main Ideas of This Section

The projection operator is a Hermitian operator that projects one vector onto the direction of another and forms a new vector in that direction; operating on a vector with the projection operators for all of the basis vectors of that space reproduces the original vector. That means that the sum of the projection operators for all the basis vectors equals the identity operator; this is a form of the completeness relation. The matrix elements of a projector operator may be found by sandwiching the operator between the bras and kets of pairs of the basis vectors or by using the outer product of the ket and bra of each basis vector.

Relevance to Quantum Mechanics
As described in Chapter 4, the projection operator is useful in determining the probability of measurement outcomes for a quantum observable by projecting the state of a system onto the eigenstates of the operator for that observable.

2.5 Expectation Values

The great quantum physicist Niels Bohr was apparently fond of a Danish proverb that says "It is difficult to predict, especially the future." Fortunately, if you've worked through the previous sections, you possess the tools to make very specific predictions about the results of measurements of quantum observables such as position, momentum, and energy.

Students new to quantum theory are often surprised to learn that such predictions can be made – after all, isn't quantum mechanics probabilistic by its very nature? So in general the results of individual measurements cannot be precisely predicted. And yet in this section you'll learn how to make very specific predictions about *average* measurement outcomes provided you know two things: the operator ($\widehat{O}$) corresponding to the measurement you plan to make, and the state of the system represented by ket $|\psi\rangle$ prior to the measurement.

Those predictions come in the form of the "expectation value" of an observable, the precise meaning of which is explained in this section. You can use the following equation to determine the expectation value of an observable (o) represented by the operator $\widehat{O}$ for a system in state $|\psi\rangle$:

$$\langle o \rangle = \langle \psi | \, \widehat{O} \, | \psi \rangle \, . \tag{2.55}$$

In this equation, the angle brackets on the left side signify the expectation value – that is, the average value of the outcome of a number of measurements of the observable associated with the operator $\widehat{O}$.

It's very important for you to understand that the phrase "number of measurements" does not refer to a sequence of observations made one after another. Instead, these measurements are made not on a single system, but on group of systems (usually called an "ensemble" of systems) all prepared to be in the same state prior to the measurement. So the expectation value is an

average over many systems, not an average over time (and it is certainly not the value you expect to get from a single measurement).

If that sounds unusual, think of the average score of all the soccer matches played on a given day. The winning sides might have an average of 2.4 goals and the losing sides an average of 1.7 goals, but would you ever expect to see a final score of 2.4 to 1.7? Clearly not, because in an individual match each side scores an integer number of goals. Only when you average over multiple matches can you expect to see non-integer values of goals scored.

This soccer match analogy is helpful in understanding why the expectation value is not the value you expect to get from an individual measurement, but it lacks one feature that's present in all quantum-mechanical observations. That feature is probability, which is the reason most quantum texts use examples such as thrown dice when introducing the concept of the expectation value. So instead of thinking about averaging a set of scores from completed matches, consider the way you might determine the expected value for the average number of goals scored by the winning side over a large number of matches if you're given a set of probabilities. For example, you might be told that the probability of the winning side scoring zero goals or more than six goals is negligible, and the probabilities of scoring one to six goals are shown in this table:

Winning side total goals	0	1	2	3	4	5	6
Probability (%)	0	22	43	18	9	5	3

Given this information, the expected number of goals ($\langle g \rangle$) for the winning team can be determined simply by multiplying each possible score (call it λ_n) by its probability (P_n) and summing the result over all possible scores:

$$\langle g \rangle = \sum_{n=1}^{N} \lambda_n P_n, \tag{2.56}$$

so in this case

$$\begin{aligned}\langle g \rangle &= \lambda_0 P_0 + \lambda_1 P_1 + \lambda_2 P_2 + \cdots + \lambda_6 P_6 \\ &= 0(0) + 1(0.22) + 2(0.43) + 3(0.18) + 4(0.09) + 5(0.05) + 6(0.03) \\ &= 2.4.\end{aligned}$$

To use this approach, you must know all of the possible outcomes and the probability of each outcome.

This same technique of multiplying each possible outcome by its probability to determine the expectation value can be used in quantum mechanics. To see how that works, consider a Hermitian operator $\widehat{O}$ and normalized wavefunction represented by the ket $|\psi\rangle$. As explained in Section 1.6, this ket can be written as the weighted combination of the kets representing the eigenvectors of operator $\widehat{O}$:

$$|\psi\rangle = c_1 |\psi_1\rangle + c_2 |\psi_2\rangle + \cdots + c_N |\psi_N\rangle = \sum_{n=1}^{N} c_n |\psi_n\rangle , \tag{1.35}$$

in which c_1 through c_N represent the amount of each orthonormal eigenfunction $|\psi_n\rangle$ in $|\psi\rangle$. Now consider the expression

$$\langle\psi| \widehat{O} |\psi\rangle ,$$

which, as described previously, can be represented as the inner product of $|\psi\rangle$ with the result of applying operator $\widehat{O}$ to ket $|\psi\rangle$. Applying the operator $\widehat{O}$ to $|\psi\rangle$ as given by Eq. 1.35 yields

$$\widehat{O} |\psi\rangle = \widehat{O} \sum_{n=1}^{N} c_n |\psi_n\rangle = \sum_{n=1}^{N} c_n \widehat{O} |\psi_n\rangle = \sum_{n=1}^{N} \lambda_n c_n |\psi_n\rangle , \tag{2.57}$$

in which λ_n represents the eigenvalue of operator $\widehat{O}$ applied to eigenket $|\psi_n\rangle$.

Now find the inner product of $|\psi\rangle$ with this expression for $\widehat{O} |\psi\rangle$. The bra $\langle\psi|$ corresponding to $|\psi\rangle$ is

$$\langle\psi| = \langle\psi_1| c_1^* + \langle\psi_2| c_2^* + \cdots + \langle\psi_N| c_N^* = \sum_{m=1}^{N} \langle\psi_m| c_m^*,$$

in which the index m is used to differentiate this summation from the summation of Eq. 2.57. This means that the inner product $(|\psi\rangle , \widehat{O} |\psi\rangle$ is

$$\begin{aligned} \langle\psi| \widehat{O} |\psi\rangle &= \sum_{m=1}^{N} \langle\psi_m| c_m^* \sum_{n=1}^{N} \lambda_n c_n |\psi_n\rangle \\ &= \sum_{m=1}^{N} \sum_{n=1}^{N} c_m^* \lambda_n c_n \langle\psi_m|\psi_n\rangle . \end{aligned}$$

But if the eigenfunctions ψ_n are orthonormal, only the terms with $n = m$ survive, so this becomes

$$\langle\psi| \widehat{O} |\psi\rangle = \sum_{n=1}^{N} \lambda_n c_n^* c_n = \sum_{n=1}^{N} \lambda_n |c_n|^2 = \langle o\rangle . \tag{2.58}$$

This has the same form as Eq. 2.56, with $|c_n|^2$ in place of the probability P_n. So the expression $\langle\psi|\,\widehat{O}\,|\psi\rangle$ will produce the expectation value $\langle o\rangle$ as long as the square magnitude of c_n represents the probability of obtaining result λ_n. As you'll see in Chapter 4, that's exactly what $|c_n|^2$ represents.

The expressions for the expectation values presented in this section can be extended to apply to situations in which the outcomes may be represented by a continuous variable x rather than discrete values λ_n. In such situations, the discrete probabilities P_n for each outcome are replaced by the continuous probability density function $P(x)$, and the sum becomes an integral over infinitesimal increments dx. The expectation value of the observable x is then

$$\langle x\rangle = \int_{-\infty}^{\infty} xP(x)dx. \tag{2.59}$$

Using the inner product, the expectation value can be written in Dirac notation and integral form as

$$\langle x\rangle = \langle\psi|\,\widehat{X}\,|\psi\rangle = \int_{-\infty}^{\infty} [\psi(x)]^*\widehat{X}[\psi(x)]dx, \tag{2.60}$$

in which $\widehat{X}$ represents the operator associated with observable x.

In quantum mechanics, expectation values play an important role in the determination of the uncertainty of a quantity such as position, momentum or energy. Calling the uncertainty in position Δx, the square of the uncertainty is given by

$$(\Delta x)^2 = \left\langle x^2\right\rangle - \langle x\rangle^2, \tag{2.61}$$

in which $\left\langle x^2\right\rangle$ represents the expectation value of the square of position (x^2) and $\langle x\rangle^2$ represents the square of the expectation value of x.

Taking the square root of both sides of Eq. 2.61 gives

$$\Delta x = \sqrt{\left\langle x^2\right\rangle - \langle x\rangle^2}. \tag{2.62}$$

As you can see on the book's website, for a distribution of position values x, $(\Delta x)^2$ is equivalent to the variance of x. That variance is defined as the average of the square of the difference between each value of x and the average value of x (that average is the expectation value $\langle x\rangle$):

$$\text{Variance of } x = (\Delta x)^2 \equiv \left\langle (x - \langle x\rangle)^2\right\rangle, \tag{2.63}$$

which means that Δx, the square root of the variance, is the standard deviation of the distribution x. Thus the uncertainty in position x may be determined

using the expectation value of the square of x and the expectation value of x in Eq. 2.62. Similarly, the uncertainty in momentum Δp is given by

$$\Delta p = \sqrt{\langle p^2 \rangle - \langle p \rangle^2}, \tag{2.64}$$

and the uncertainty in energy ΔE may be found using

$$\Delta E = \sqrt{\langle E^2 \rangle - \langle E \rangle^2}. \tag{2.65}$$

Main Idea of This Section
The expression $\langle \psi \vert \widehat{O} \vert \psi \rangle$ gives the expectation value of the observable associated with operator $\widehat{O}$ for a system in quantum state $\vert \psi \rangle$.

Relevance to Quantum Mechanics
When Schrödinger published his equation in 1926, the meaning of the wavefunction ψ became the subject of debate. Later that year, German physicist Max Born published a paper in which he related the solutions of the Schrödinger equation to the probability of measurement outcomes, stating in a footnote that "A more precise consideration shows that the probability is proportional to the square" of the quantities we've called c_n in Eq. 2.58. You can read more about the "Born rule" in Chapter 4.

2.6 Problems

1. For vector $\vec{A} = A_x \hat{\imath} + A_y \hat{\jmath}$ and matrix $\bar{\bar{R}} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$, what does operator $\bar{\bar{R}}$ do to vector $\vec{A}$? (Hint: consider $\bar{\bar{R}}\vec{A}$ for the cases $\theta = 90^\circ$ and $\theta = 180^\circ$.)

2. Show that the complex vectors $\begin{pmatrix} 1 \\ i \end{pmatrix}$ and $\begin{pmatrix} 1 \\ -i \end{pmatrix}$ are eigenvectors of matrix $\bar{\bar{R}}$ in Problem 1, and find the eigenvalues of each eigenvector.

3. The discussion around Eq. 2.8 shows that $\sin(kx)$ is not an eigenfunction of the spatial first-derivative operator d/dx. Is $\cos(kx)$ an eigenvector of that operator? What about $\cos(kx) + i\sin(kx)$ or $\cos(kx) - i\sin(kx)$? If so, find the eigenvalues for these eigenvectors.

4. If operator $\widehat{M}$ has matrix representation $\bar{\bar{M}} = \begin{pmatrix} 2 & 1+i \\ 1-i & 3 \end{pmatrix}$ in 2-D Cartesian coordinates,

a) Show that $\begin{pmatrix} 1+i \\ -1 \end{pmatrix}$ and $\begin{pmatrix} \frac{1+i}{2} \\ 1 \end{pmatrix}$ are eigenvectors of $\widehat{M}$.

b) Normalize these eigenvectors and show that they're orthogonal.

c) Find the eigenvalues for these eigenvectors.

d) Find the matrix representation of operator $\widehat{M}$ in the basis system of these eigenvectors.

5. Consider the matrices $\bar{\bar{A}} = \begin{pmatrix} 5 & 0 \\ 0 & i \end{pmatrix}$ and $\bar{\bar{B}} = \begin{pmatrix} 3+i & 0 \\ 0 & 2 \end{pmatrix}$.

a) Do these matrices commute?

b) Do matrices $\bar{\bar{C}} = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$ and $\bar{\bar{D}} = \begin{pmatrix} c & 0 \\ 0 & d \end{pmatrix}$ commute?

c) For matrices $\bar{\bar{E}} = \begin{pmatrix} 2 & i \\ 3 & 5i \end{pmatrix}$ and $\bar{\bar{F}} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ find the relationships between a, b, c, and d that ensure that $\bar{\bar{E}}$ and $\bar{\bar{F}}$ commute.

6. Specify whether each of the following matrices are Hermitian (for parts d through f, fill in the missing elements to make these matrices Hermitian, if possible):

a) $\bar{\bar{A}} = \begin{pmatrix} 5 & 1 \\ 1 & 2 \end{pmatrix}$ b) $\bar{\bar{B}} = \begin{pmatrix} i & -3i \\ 3i & 0 \end{pmatrix}$ c) $\bar{\bar{C}} = \begin{pmatrix} 2 & 1+i \\ 1-i & 3 \end{pmatrix}$

d) $\bar{\bar{D}} = \begin{pmatrix} 0 & \frac{i}{2} \\ & 4 \end{pmatrix}$ e) $\bar{\bar{E}} = \begin{pmatrix} i & 3 \\ 3 & \end{pmatrix}$ f) $\bar{\bar{F}} = \begin{pmatrix} 2 & \\ 5i & 1 \end{pmatrix}$.

7. Find the elements of the matrices representing the projection operators $\widehat{P}_1$, $\widehat{P}_2$, and $\widehat{P}_3$ in the coordinate system with orthogonal basis vectors $\vec{\epsilon}_1 = 4\hat{\imath} - 2\hat{\jmath}$, $\vec{\epsilon}_2 = 3\hat{\imath} + 6\hat{\jmath}$, and $\vec{\epsilon}_3 = \hat{k}$.

8. Use the projection operators from Problem 7 to project vector $\vec{A} = 7\hat{\imath} - 3\hat{\jmath} + 2\hat{k}$ onto the directions of $\vec{\epsilon}_1$, $\vec{\epsilon}_2$, and $\vec{\epsilon}_3$.

9. Consider a six-sided die labeled with numbers 1 through 6.

a) If the die is fair, the probability of occurrence of any number (1 through 6) is equal. Find the expectation value and standard deviation in this case.

b) If the die is "loaded," the probability of occurrence might be:

Number	1	2	3	4	5	6
Probability (%)	10	70	15	3	1	1

What are the expectation value and standard deviation in this case?

10. Operating on the orthonormal basis kets $|\epsilon_1\rangle$, $|\epsilon_2\rangle$, and $|\epsilon_3\rangle$ with operator $\widehat{O}$ produces the results $\widehat{O}\,|\epsilon_1\rangle = 2\,|\epsilon_1\rangle$, $\widehat{O}\,|\epsilon_2\rangle = -i\,|\epsilon_1\rangle + |\epsilon_2\rangle$, and $\widehat{O}\,|\epsilon_3\rangle = |\epsilon_3\rangle$. If $\psi = 4\,|\epsilon_1\rangle + 2\,|\epsilon_2\rangle + 3\,|\epsilon_3\rangle$, what is the expectation value $\langle o\rangle$?

3

The Schrödinger Equation

If you've worked through Chapters 1 and 2, you've already seen several references to the **Schrödinger equation** and its solutions. As you'll learn in this chapter, the Schrödinger equation describes how a quantum state evolves over time, and understanding the physical meaning of the terms of this powerful equation will prepare you to understand the behavior of quantum wavefunctions. So this chapter is all about the Schrödinger equation, and you can read about the solutions to the Schrödinger equation in Chapters 4 and 5.

In the first section of this chapter, you'll see a "derivation" of several forms of the Schrödinger equation, and you'll learn why the word "derivation" is in quotes. Then, in Section 3.2, you'll find a description of the meaning of each term in the Schrödinger equation as well as an explanation of exactly what the Schrödinger equation tells you about the behavior of quantum wavefunctions. The subject of Section 3.3 is a time-independent version of the Schrödinger equation that you're sure to encounter if you read more advanced quantum books or take a course in quantum mechanics.

To help you focus on the physics of the situation without getting too bogged down in mathematical notation, the Schrödinger equation discussed in most of this chapter is a function of only one spatial variable (x). As you'll see in later chapters, even this one-dimensional treatment will let you solve several interesting problems in quantum mechanics, but for certain situations you're going to need the three-dimensional version of the Schrödinger equation. So that's the subject of the final section of this chapter (Section 3.4).

3.1 Origin of the Schrödinger Equation

If you look at the introduction of the Schrödinger equation in popular quantum texts, you'll find that there are several ways to "derive" the Schrödinger equation. But as the authors of those texts invariably point out, none of those methods are rigorous derivations from first principles (hence the quotation marks). As the brilliant and always-entertaining physicist Richard Feynman said, "It's not possible to derive it from anything you know. It came out of the mind of Schrödinger."

So if Erwin Schrödinger didn't arrive at this equation from first principles, how exactly did he get there? The answer is that although his approach evolved over several papers, from the start Schrödinger clearly recognized the need for a wave equation from the work of French physicist Louis de Broglie. But Schrödinger also realized that unlike the classical wave equation, which is a second-order partial differential equation in both space and time, the form of the quantum wave equation should be first-order in time, for reasons explained later in this chapter. Importantly, he also saw that making the equation complex (that is, including a factor of $\sqrt{-1}$ in one of the coefficients) provided immense benefits.

One approach to understanding the basis of the Schrödinger equation is to begin with the classical equation relating total energy to the sum of kinetic energy and potential energy. To apply this principle to quantum wavefunctions[1], begin with the equation developed by Max Planck and Albert Einstein in the early twentieth century relating the energy of a photon (E) to its frequency (f) or angular frequency ($\omega = 2\pi f$):

$$E = hf = \hbar\omega, \tag{3.1}$$

in which h represents the **Planck constant** and $\hbar$ is the modified Planck constant ($\hbar = \frac{h}{2\pi}$).

Another useful equation comes from James Clerk Maxwell's work on **radiation pressure** in 1862 and his determination that electromagnetic waves carry momentum. The magnitude of that momentum (p) is related to energy (E) and the speed of light (c) by the relation

$$p = \frac{E}{c}. \tag{3.2}$$

In 1924, de Broglie suggested that at the quantum level particles can exhibit wavelike behavior, and the momentum of these "matter waves" can be

[1] You can read about the relationship of quantum wavefunctions to quantum states in Section 4.2.

determined by combining the **Planck–Einstein relation** ($E = \hbar\omega$) with the momentum-energy relation

$$p = \frac{E}{c} = \frac{\hbar\omega}{c}. \tag{3.3}$$

Since the frequency (f) of a wave is related to its wavelength (λ) and speed (c) by the equation $f = \frac{c}{\lambda}$, the momentum may be written as

$$\begin{aligned} p &= \frac{\hbar\omega}{c} = \frac{\hbar(2\pi f)}{c} = \frac{\hbar\left(2\pi\frac{c}{\lambda}\right)}{c} \\ &= \frac{\hbar 2\pi}{\lambda}. \end{aligned}$$

The definition of wavenumber ($k \equiv \frac{2\pi}{\lambda}$) makes this

$$p = \hbar k. \tag{3.4}$$

This equation is known as de Broglie's relation, and it represents the mixing of wave and particle behavior into the concept of "**wave-particle duality**."

Since momentum is the product of mass and velocity ($p = mv$) in the nonrelativistic case, the classical equation for kinetic energy is

$$KE = \frac{1}{2}mv^2 = \frac{p^2}{2m} \tag{3.5}$$

and substituting $\hbar k$ for momentum (Eq. 3.4) gives

$$KE = \frac{\hbar^2 k^2}{2m}. \tag{3.6}$$

Now write the total energy (E) as the sum of the kinetic energy (KE) plus the potential energy (V):

$$E = KE + V = \frac{\hbar^2 k^2}{2m} + V, \tag{3.7}$$

and since $E = \hbar\omega$ (Eq. 3.1), the total energy is given by

$$E = \hbar\omega = \frac{\hbar^2 k^2}{2m} + V. \tag{3.8}$$

This equation provides the foundation for the Schrödinger equation when applied to a quantum wavefunction $\Psi(x, t)$.

To get from Eq. 3.8 to the Schrödinger equation, one path is to assume that the quantum wavefunction has the form of a wave for which the surfaces of

constant phase are flat planes.[2] For a **plane wave** propagating in the positive x-direction, the wavefunction is given by

$$\Psi(x,t) = Ae^{i(kx-\omega t)}, \tag{3.9}$$

in which A represents the wave's amplitude, k is the wavenumber, and ω is the angular frequency of the wave.

With this expression for Ψ, taking temporal and spatial derivatives is straightforward (and helpful in getting from Eq. 3.8 to the Schrödinger equation). Starting with the first partial derivative of $\Psi(x,t)$ with respect to time (t),

$$\frac{\partial \Psi(x,t)}{\partial t} = \frac{\partial \left[Ae^{i(kx-\omega t)}\right]}{\partial t} = -i\omega\left[Ae^{i(kx-\omega t)}\right] = -i\omega\Psi(x,t). \tag{3.10}$$

So for the plane-wave function of Eq. 3.9, taking the first partial derivative with respect to time has the effect of returning the original wavefunction multiplied by $-i\omega$:

$$\frac{\partial \Psi}{\partial t} = -i\omega\Psi, \tag{3.11}$$

which means that you can write ω as

$$\omega = \frac{1}{-i\Psi}\frac{\partial \Psi}{\partial t} = i\frac{1}{\Psi}\frac{\partial \Psi}{\partial t}, \tag{3.12}$$

in which the relation $\frac{1}{i} = \frac{-(i)(i)}{i} = -i$ has been used.

Now consider what happens when you take the first partial derivative of $\Psi(x,t)$ with respect to space (x in this case):

$$\frac{\partial \Psi(x,t)}{\partial x} = \frac{\partial \left[Ae^{i(kx-\omega t)}\right]}{\partial x} = ik\left[Ae^{i(kx-\omega t)}\right] = ik\Psi(x,t). \tag{3.13}$$

So taking the first partial derivative of the plane-wave function with respect to distance (x) has the effect of returning the original wavefunction multiplied by ik:

$$\frac{\partial \Psi}{\partial x} = ik\Psi. \tag{3.14}$$

It's also helpful to note the effect of taking the second partial spatial derivative of the plane-wave function, which gives

$$\frac{\partial^2 \Psi(x,t)}{\partial x^2} = \frac{\partial \left[ikAe^{i(kx-\omega t)}\right]}{\partial x} = ik\left[ikAe^{i(kx-\omega t)}\right] = -k^2\Psi(x,t), \tag{3.15}$$

[2] If you're unfamiliar with plane waves, you can see a sketch of the planes of constant phase in Fig. 3.4 in Section 3.4.

which means that taking the second partial derivative with respect to x has the effect of returning the original wavefunction multiplied by $-k^2$:

$$\frac{\partial^2 \Psi}{\partial x^2} = -k^2 \Psi. \tag{3.16}$$

So just as the angular frequency ω may be written in terms of the wavefunction Ψ and its temporal partial derivative $\frac{\partial \Psi}{\partial t}$ (Eq. 3.12), the square of the wavenumber k may be written in term of Ψ and its second spatial partial derivative $\frac{\partial^2 \Psi}{\partial x^2}$:

$$k^2 = -\frac{1}{\Psi}\frac{\partial^2 \Psi}{\partial x^2}. \tag{3.17}$$

What good has it done to write ω and k^2 in terms of the wavefunction Ψ and its derivatives? To understand that, look back at Eq. 3.8, and note that it includes a factor of ω on the left side and a factor of k^2 on the right side of the second equals sign. Substituting the expression for ω from Eq. 3.12 into the left side gives

$$E = \hbar\omega = \hbar\left(i\frac{1}{\Psi}\frac{\partial \Psi}{\partial t}\right) = i\hbar\frac{1}{\Psi}\frac{\partial \Psi}{\partial t}. \tag{3.18}$$

Likewise, substituting the expression for k^2 from Eq. 3.17 into the right side of Eq. 3.8 gives

$$\frac{\hbar^2 k^2}{2m} + V = \frac{\hbar^2}{2m}\left(-\frac{1}{\Psi}\frac{\partial^2 \Psi}{\partial x^2}\right) + V, \tag{3.19}$$

which makes the equation for total energy look like this:

$$i\hbar\frac{1}{\Psi}\frac{\partial[\Psi(x,t)]}{\partial t} = -\frac{\hbar^2}{2m}\frac{1}{\Psi}\frac{\partial^2[\Psi(x,t)]}{\partial x^2} + V, \tag{3.20}$$

and multiplying through by the wavefunction $\Psi(x,t)$ yields

$$i\hbar\frac{\partial[\Psi(x,t)]}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2[\Psi(x,t)]}{\partial x^2} + V[\Psi(x,t)]. \tag{3.21}$$

This is the most common form of the one-dimensional time-dependent Schrödinger equation. The physical meaning of this equation and each of its terms is discussed in this chapter, but before getting to that, you should consider how we got here. Writing the total energy as the sum of the kinetic energy and the potential energy is perfectly general, but to get to Eq. 3.21, we used the expression for a plane wave. Specifically, Eq. 3.12 for ω and Eq. 3.17 for k^2 resulted from the temporal and spatial derivatives of the plane-wave

function (Eq. 3.9). Why should we expect this equation to hold for quantum wavefunctions of other forms?

One answer is this: it works. That is, wavefunctions that are solutions to the Schrödinger equation lead to predictions that agree with laboratory measurements of quantum observables such as position, momentum, and energy.

If it seems surprising that an equation based on a simple plane-wave function describes the behavior of particles and systems that have little in common with plane waves, note that the Schrödinger equation is linear, which means that the terms involving the wavefunction, such as $\frac{\partial[\Psi(x,t)]}{\partial t}$, $\frac{\partial^2[\Psi(x,t)]}{\partial x^2}$, and $V\Psi(x,t)$, are all raised to the first power.[3] As you may recall, a linear equation has the supremely useful characteristic that *superposition* works, which guarantees that combinations of solutions are also solutions. And since plane waves are solutions to the Schrödinger equation, the linear nature of the equation means that superpositions of plane waves are also solutions. By judicious combination of plane waves, a variety of quantum wavefunctions may be synthesized, just as a variety of functions may be synthesized from the sine and cosine functions in Fourier analysis.

To understand why that works, consider the wavefunction of a quantum particle that is localized over some region of the x-axis. Since a single-frequency plane wave extends to infinity in both directions (±x), it's clear that additional frequency components are needed to restrict the particle's wavefunction to the desired region. Combining those components in just the right proportion allows you to form a "wave packet" with amplitude that rolls off with distance from the center of the packet.

To form a wavefunction from a finite number (N) of discrete plane-wave components, a weighted linear combination may be used:

$$\begin{aligned}\Psi(x,t) &= A_1 e^{i(k_1 x-\omega_1 t)} + A_2 e^{i(k_2 x-\omega_2 t)} + \cdots + A_N e^{i(k_N x-\omega_N t)} \\ &= \Sigma_{n=1}^{N} A_n e^{i(k_n x-\omega_n t)},\end{aligned} \tag{3.22}$$

in which A_n, k_n, and ω_n represent the amplitude, wave number, and angular frequency of the nth plane-wave component, respectively. Note that the constants A_n determine the "amount" of each plane wave included in the mix.

[3] Remember that the second-order derivative $\frac{\partial^2\Psi}{\partial x^2}$ represents the change in the slope of Ψ with respect to x, which is not the same as the square of the slope $(\frac{\partial\Psi}{\partial x})^2$. So $\frac{\partial^2\Psi}{\partial x^2}$ is a second-order derivative, but it's raised to the first power in the Schrödinger equation.

Alternatively, a wavefunction satisfying the Schrödinger equation can be synthesized using a continuous spectrum of plane waves:

$$\Psi(x,t) = \int_{-\infty}^{\infty} A(k)e^{i(kx-\omega t)}dk, \tag{3.23}$$

in which the summation of Eq. 3.22 is now an integral and the discrete amplitudes A_n have been replaced by a continuous function of wavenumber $A(k)$. As in the discrete case, this function is related to the amplitude of the plane-wave components as a function of wavenumber. Specifically, in the continuous case $A(k)$ represents the amplitude per unit wavenumber.

And just as in the case of an individual plane wave, taking the first-order time derivative and second-order spatial derivative of wavefunctions synthesized from combinations of plane waves leads to the Schrödinger equation.

A very common and useful version of Eq. 3.23 can be obtained by pulling a constant factor of $1/\sqrt{2\pi}$ out of the weighting function $A(k)$ and setting the time to an initial reference time ($t = 0$):

$$\psi(x) = \Psi(x,0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \phi(k)e^{ikx}dk. \tag{3.24}$$

This version makes clear the Fourier-transform relationship between the position-based wavefunction $\psi(x)$ and the wavenumber-based wavefunction $\phi(k)$, which plays an important role in Chapters 4 and 5. You can read about Fourier transforms in Section 4.4.

Before considering exactly what the Schrödinger equation tells you about the behavior of quantum wavefunctions, it's worthwhile to consider another form of the Schrödinger equation that you're very likely to encounter in textbooks on quantum mechanics. That version of Eq. 3.21 looks like this:

$$i\hbar\frac{\partial\Psi}{\partial t} = \widehat{H}\Psi. \tag{3.25}$$

In this equation, $\widehat{H}$ represents the "Hamiltonian," or total-energy operator. Equating the right sides of this equation and Eq. 3.21 gives

$$\widehat{H}\Psi = -\frac{\hbar^2}{2m}\frac{\partial^2\Psi}{\partial x^2} + V\Psi,$$

which means that the **Hamiltonian operator** is equivalent to

$$\widehat{H} \equiv -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V. \tag{3.26}$$

To see why this makes sense, use the relations $p = \hbar k$ and $E = \hbar\omega$ to rewrite the plane-wave function in terms of momentum (p) and energy (E)

$$\begin{aligned}\Psi(x,t) &= Ae^{i(kx-\omega t)} = Ae^{i\left(\frac{p}{\hbar}x-\frac{E}{\hbar}t\right)} \\ &= Ae^{\frac{i}{\hbar}(px-Et)},\end{aligned} \tag{3.27}$$

and then take the first-order spatial derivative:

$$\frac{\partial\Psi}{\partial x} = \left(\frac{i}{\hbar}p\right)Ae^{\frac{i}{\hbar}(px-Et)} = \left(\frac{i}{\hbar}p\right)\Psi$$

or

$$p\Psi = \frac{\hbar}{i}\frac{\partial\Psi}{\partial x} = -i\hbar\frac{\partial\Psi}{\partial x}. \tag{3.28}$$

This suggests that the (one-dimensional) differential operator associated with momentum may be written as

$$\widehat{p} = -i\hbar\frac{\partial}{\partial x}. \tag{3.29}$$

This is a very useful relation in its own right, but for now you can use it to justify Eq. 3.26 for the Hamiltonian operator. To do that, write an operator version of the classical total-energy equation $E = \frac{p^2}{2m} + V$:

$$\widehat{H} = \frac{(\widehat{p})^2}{2m} + V = \frac{\left(-i\hbar\frac{\partial}{\partial x}\right)^2}{2m} + V, \tag{3.30}$$

in which $\widehat{H}$ is an operator associated with the total energy E.

Now recall that, unlike the square of an algebraic quantity, the square of an operator is formed by applying the operator twice. For example, the square of operator $\widehat{O}$ operating on function Ψ is

$$(\widehat{O})^2\Psi = \widehat{O}(\widehat{O}\Psi),$$

so

$$\begin{aligned}(\widehat{p})^2\Psi = \widehat{p}(\widehat{p}\Psi) &= -i\hbar\frac{\partial}{\partial x}\left(-i\hbar\frac{\partial\Psi}{\partial x}\right) \\ &= i^2\hbar^2\frac{\partial^2\Psi}{\partial x^2} = -\hbar^2\frac{\partial^2\Psi}{\partial x^2}.\end{aligned} \tag{3.31}$$

Thus the $(\widehat{p})^2$ operator may be written as

$$(\widehat{p})^2 = -\hbar^2\frac{\partial^2}{\partial x^2} \tag{3.32}$$

and plugging this expression into Eq. 3.30 gives

$$\begin{aligned}\widehat{H} &= \frac{(\widehat{p})^2}{2m} + V = -\frac{\hbar^2 \frac{\partial^2}{\partial x^2}}{2m} + V \\ &= \frac{-\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V,\end{aligned}$$

in agreement with Eq. 3.26.

In the next section, you can read more about the meaning of each term in the Schrödinger equation as well as the meaning of the equation as a whole. And if you'd like to see some alternative approaches to "deriving" the Schrödinger equation, on the book's website you can find descriptions of the "probability flow" approach and the "path integral" approach along with links to helpful websites for those approaches.

3.2 What the Schrödinger Equation Means

Once you understand where the Schrödinger equation comes from, it's worthwhile to step back to ask "What is this equation telling me?" To help you understand the answer to that question, Fig. 3.1 shows an expanded view of the Schrödinger equation in which each term is defined, followed by a brief description of each term and the dimensions and SI units of each term:

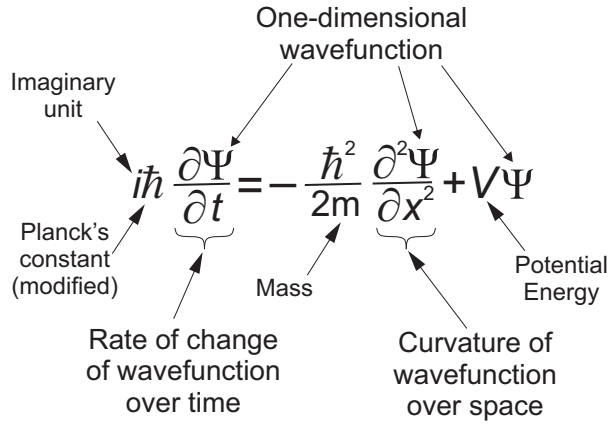

One-dimensional wavefunction
Imaginary unit
$i\hbar \frac{\partial \Psi}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2 \Psi}{\partial x^2} + V\Psi$
Planck's constant (modified)
Mass
Potential Energy
Rate of change of wavefunction over time
Curvature of wavefunction over space

Figure 3.1 Expanded view of Schrödinger equation.

$\frac{\partial \Psi}{\partial t}$: The quantum wavefunction $\Psi(x,t)$ is a function of both time and space, so this term represents the change in the wavefunction over time only (which is why it's a partial derivative). In a graph of the wavefunction at a given location as a function of time, this term is the slope of the graph. To determine the dimensions of this term, note that the one-dimensional quantum wavefunction Ψ represents a probability density amplitude (which you can read about in Chapter 4), the square of which has dimensions of probability per unit length. This is equivalent to $\frac{1}{\text{m}}$ in the SI system, since probability is dimensionless. And if Ψ^2 has units of $\frac{1}{\text{m}}$, then Ψ must have units of $\frac{1}{\sqrt{\text{m}}}$, which means that $\frac{\partial \Psi}{\partial t}$ has SI units of $\frac{1}{\text{s}\sqrt{\text{m}}}$.

i: The numerical value of the imaginary unit i is $\sqrt{-1}$, as described in Section 1.4. As an operator, multiplication by i has the effect of causing a 90° rotation in the complex plane (Fig. 1.7), moving numbers from the positive real axis to the positive imaginary axis, or from the positive imaginary axis to the negative real axis, for example. The presence of i in the Schrödinger equation means that the quantum wavefunction solutions may be complex, and this significantly impacts the result of combining wavefunctions, as you can see in Chapters 4 and 5. The factor i is dimensionless.

$\hbar$: The modified Planck constant $\hbar$ is the Planck constant h divided by 2π. Just as h is the constant of proportionality between the energy (E) and frequency (f) of a photon ($E = hf$), $\hbar$ is the constant of proportionality between total energy (E) and angular frequency (ω), and between momentum (p) and wavenumber (k) in quantum wavefunctions, as shown in the equations $E = \hbar\omega$ and $p = \hbar k$.

These two equations account for the presence of the modified Planck constant in the Schrödinger equation. The modified Planck constant $\hbar$ appears in the numerator of the factor multiplying $\frac{\partial \Psi}{\partial t}$ on one side of the Schrödinger equation because it appears in the total-energy equation $E = \hbar\omega$, and the square of $\hbar$ appears in the numerator of the factor multiplying $\frac{\partial^2 \Psi}{\partial x^2}$ because it appears in the momentum equation $p = \hbar k$, which gives rise to the kinetic-energy expression $KE = \frac{(\hbar k)^2}{2m}$.

The Planck constant h has dimensions of energy per unit frequency, so its SI units are Joules per Hertz (equivalent to Js or $\text{m}^2\text{kg/s}$), while $\hbar$ has dimensions of Joules per Hertz per radian (equivalent to Js/rad or $\text{m}^2\text{kg/s rad}$). The numerical values of these constants in the SI system are $h = 6.62607 \times 10^{-34}$ Js and $\hbar = 1.05457 \times 10^{-34}$ Js/rad.

m: The mass of the particle or system associated with the quantum wavefunction $\Psi(x,t)$ is a measure of inertia, that is, resistance to acceleration. In the SI system, mass has units of kilograms.

$\frac{\partial^2\Psi}{\partial x^2}$: This second-derivative term represents the curvature of the wavefunction over space (that is, over x in the one-dimensional case). Since $\Psi(x,t)$ is a function of both space and time, the first partial derivative $\frac{\partial\Psi}{\partial x}$ gives the change of the wavefunction over space (the slope of the wavefunction plotted against x), and the second partial derivative $\frac{\partial^2\Psi}{\partial x^2}$ gives the change in the slope of the wavefunction over space (that is, the curvature of the wavefunction).

Since $\Psi(x,t)$ has SI units of $\frac{1}{\sqrt{\text{m}}}$, as described earlier, the term $\frac{\partial^2\Psi}{\partial x^2}$ has units of $\frac{1}{\text{m}^2\sqrt{\text{m}}} = \frac{1}{(\text{m})^{5/2}}$.

V: The potential energy of the system may vary over space and time, in which case you'll see this term written as $V(x,t)$ for the one-dimensional case or as $V(\vec{r},t)$ in the three-dimensional case. Note that some physics texts use V to denote the electrostatic potential (potential energy per unit charge, with units of Joules per Coulomb or volts), but in quantum mechanics texts the words "potential" and "potential energy" tend to be used interchangeably.

Unlike classical mechanics, in which the potential, kinetic, and total energy have precise values, and in which the potential energy cannot exceed the total energy, in quantum mechanics only the average or expectation value of the energy may be determined, and a particle's total energy may be less than the potential energy in some regions.

The behavior of quantum wavefunctions in these classically "unallowed" regions (in which $E < V$) is very different from their behavior in classically "allowed" regions (in which $E \geq V$). As you'll see in the next section of this chapter, for the "stationary solutions" of the time-independent version of the Schrödinger equation, the difference between the total energy and the potential energy determines the wavelength for oscillating solutions in classically allowed regions and the rate of decay for evanescent solutions in classically unallowed regions.

As you may have guessed, the potential-energy term in the Schrödinger equation has dimensions of energy and SI units of Joules (equivalent to kg m^2/s^2).

So the individual terms of the Schrödinger equation are readily understandable, but the real power of this equation comes from the relationship between

those terms. Taken together, the terms of the Schrödinger equation form a parabolic second-order partial differential equation. Here's why each of those terms applies:

Differential because the equation involves the *change* in the wavefunction Ψ (that is, the derivatives of $\Psi(x,t)$ over space and time);

Partial because the wavefunction $\Psi(x,t)$ depends on both space (x) and time (t);

Second-order because the highest derivative ($\frac{\partial^2\Psi}{\partial x^2}$) in the equation is a second derivative;

Parabolic because the combination of a first-order differential term ($\frac{\partial\Psi}{\partial t}$) and a second-order differential term ($\frac{\partial^2\Psi}{\partial x^2}$) is analogous to the combination of a first-order algebraic term (y) and a second-order algebraic term (x^2) in the equation of a parabola ($y = cx^2$).

These terms describe what the Schrödinger equation is, but what does it mean? To understand that, you may find it helpful to consider a well-known equation in classical physics:

$$\frac{\partial[f(x,t)]}{\partial t} = D\frac{\partial^2[f(x,t)]}{\partial x^2}. \tag{3.33}$$

This one-dimensional "diffusion" equation[4] describes the behavior of a quantity $f(x,t)$ with spatial distribution that may evolve over time, such as the concentration of a substance or the temperature of a fluid. In the diffusion equation, the proportionality factor "D" between the first-order time derivative and the second-order space derivative represents the diffusion coefficient.

To see the similarity between the classical diffusion equation and the Schrödinger equation, consider the case in which the potential energy (V) is zero, and write Eq. 3.21 as

$$\frac{\partial[\Psi(x,t)]}{\partial t} = \frac{i\hbar}{2m}\frac{\partial^2[\Psi(x,t)]}{\partial x^2}. \tag{3.34}$$

Comparing this form of the Schrödinger equation to the diffusion equation, you can see that both relate the first-order time derivative of a function to the second-order spatial derivative of that function. But as you might expect, the presence of the "i" factor in the Schrödinger equation has important implications for the wavefunctions that are solutions of that equation, and you can read about those implications in Chapters 4 and 5. But for now, you should make sure you understand the fundamental relationship in both

[4] This equation is also called the heat equation or Fick's second law.

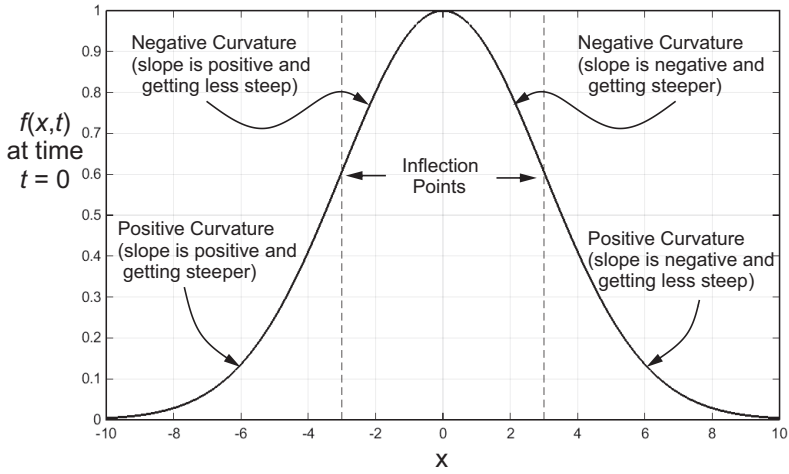


Figure 3.2 Regions of positive and negative curvature for peaked waveform.

of these equations: the evolution of the waveform over time is proportional to the curvature of the waveform over space.

And why should the rate of change of a function be related to the spatial curvature of that function? To understand that, consider the function $f(x, t)$ shown in Fig. 3.2 for time $t = 0$. This function could represent, for example, the initial temperature distribution of a fluid with a warm spot in the region of $x = 0$. To determine how this temperature distribution will evolve over time, the diffusion equation tells you to consider the curvature of the wavefunction in various regions.

As you can see in the figure, this function has a maximum at $x = 0$ and inflection points[5] at $x = -3$ and $x = +3$. For the region to the left of the inflection point at $x = -3$, the slope of the function ($\frac{\partial f}{\partial x}$) is positive and getting more positive as x increases, which means the curvature in this region (that is, the *change* in the slope $\frac{\partial^2 f}{\partial x^2}$) is positive. Likewise, to the right of the inflection point at $x = +3$, the slope of the function is negative but getting less negative with increasing x, meaning that the curvature (again, the change in the slope) is positive in this region as well.

Now consider the regions between $x = -3$ and $x = 0$ and between $x = 0$ and $x = +3$. Between $x = -3$ and $x = 0$, the slope of the function is positive

[5] An inflection point is a location at which the sign of the curvature changes.

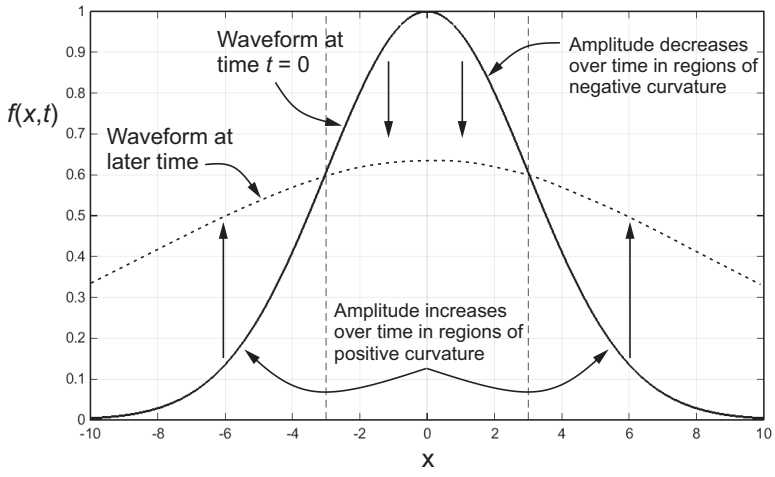


Figure 3.3 Time evolution for regions of positive and negative curvature.

but becoming less steep with increasing x, so the curvature in this region is negative. And between $x = 0$ and $x = +3$, the slope is negative and becoming steeper with increasing x, so the curvature in this region is also negative.

And here's the payoff: since the diffusion equation tells you that the time rate of change of the function $f(x,t)$ is proportional to the curvature of the function, the function will evolve as shown in Fig. 3.3.

As you can see in that figure, the function $f(x,t)$ will increase in regions of positive curvature ($x < -3$ and $x > +3$) and will decrease in regions of negative curvature ($-3 < x < +3$). If $f(x,t)$ represents temperature, for example, this is exactly what you'd expect as the energy from the initially warm region diffuses into the cooler neighboring regions.

So given the similarity between the Schrödinger equation and the classical diffusion equation, does that mean that all quantum particles and systems will somehow "diffuse" or spread out in space as time passes? If so, exactly what is it that's spreading out?

The answer to the first of these questions is "Sometimes, but not always." The reason for that answer can be understood by considering an important difference between the Schrödinger equation and the diffusion equation. That difference is the factor of "i" in the Schrödinger equation, which means that the wavefunction (Ψ) can be complex. And as you'll see in Chapters 4 and 5, complex wavefunctions may exhibit wavelike (oscillatory) behavior rather than diffusing under some circumstances.

As for the question about what's spreading out (or oscillating), for that answer we turn to Max Born, whose 1926 interpretation of the wavefunction Ψ as a probability amplitude is now widely accepted and is a fundamental precept of the Copenhagen interpretation of quantum mechanics, which you can read about in Chapter 4. According to the "Born rule," the modulus squared ($|\Psi|^2 = \Psi^*\Psi$) of a particle's position-space wavefunction gives the particle's position probability density function (that is, the probability per unit length in the one-dimensional case). This means that the integral of the position probability density function over any spatial interval gives the probability of finding the particle within that interval. So when the wavefunction oscillates or diffuses, it's the probability distribution that's changing.

Here's another propitious characteristic of the Schrödinger equation: the time derivative $\frac{\partial \Psi}{\partial t}$ is first-order, which differs from the second-order time derivative of the classical wave equation. Why is that helpful? Because a first-order time derivative tells you how fast the wavefunction itself is changing over time, which means that knowledge of the wavefunction at some instant in time completely specifies the state of the particle or system at all future times. That's consistent with the principle that the wavefunctions that satisfy the Schrödinger equation represent "all you can ever know" about the state of a particle or system.

But if you're a fan of the classical wave equation with its second-order time and spatial derivatives, you may be wondering whether it's useful to take another time derivative of the Schrödinger equation. That's certainly possible, but recall that taking another time derivative would pull down another factor of ω from the plane-wave function $e^{i(kx-\omega t)}$, and ω is proportional to E by de Broglie's relation ($E = \hbar\omega$). That means that the resulting equation would include the particle's energy as a coefficient of the time-derivative term.

You may be thinking, "But don't all equations of motion depend on energy?" Definitely not, as you can see by considering Newton's Second Law: $\vec{F} = m\vec{a}$, better written as $\vec{a} = \Sigma\vec{F}/m$. This says that the acceleration of an object is directly proportional to the vector sum of the forces acting on it, and inversely proportional to the object's mass. But in classical physics the acceleration does not depend on the energy, momentum, or velocity of the object. So if the Schrödinger equation is to serve a purpose in quantum mechanics similar to that of Newton's Second Law in classical mechanics, the time evolution of the wavefunction shouldn't depend on the particle or system's energy or momentum. Hence the time derivative cannot be of second order.

So although the Schrödinger equation can't be derived from first principles, the form of the equation does make sense. More importantly, it gives results

that predict and describe the behavior of quantum particles and systems over space and time. But one very useful form of the Schrödinger equation is independent of time, and that version is the subject of the next section.

3.3 Time-Independent Schrödinger Equation

Separating out the time-dependent and space-dependent terms of the Schrödinger equation is helpful in understanding why the quantum wavefunction behaves as it does. That separation can be accomplished, as for many differential equations, using the technique of separation of variables.

This technique begins with the assumption that the solution ($\Psi(x,t)$ in this case) can be written as the product of two separate functions, one depending only on x and the other depending only on t. You may have encountered this technique in one of your physics or mathematics classes, and you may recall that there's no a priori reason why this approach should work. But it often does work, and in any situation in which the potential energy varies only over space (and not over time), you can use separation of variables to solve the Schrödinger equation.

To see how this works, start by writing the quantum wavefunction as the product of the function $\psi(x)$ (which depends only on space) and the function $T(t)$ (which depends only on time):

$$\Psi(x,t) = \psi(x)T(t). \tag{3.35}$$

Inserting this into the Schrödinger equation gives

$$i\hbar\frac{\partial[\psi(x)T(t)]}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2[\psi(x)T(t)]}{\partial x^2} + V[\psi(x)T(t)].$$

And here's the reason that separation of variables is powerful: since the function $\psi(x)$ depends only on location (x) and not on time (t), you can pull the $\psi(x)$ term out of the partial derivative with respect to t. Likewise, since the function $T(t)$ depends only on time and not on location, you can pull the $T(t)$ term out of the second partial derivative with respect to x. Doing this gives

$$i\hbar\psi(x)\frac{d[T(t)]}{dt} = -\frac{\hbar^2 T(t)}{2m}\frac{d^2[\psi(x)]}{dx^2} + V[\psi(x)T(t)],$$

in which the partial derivatives have become full derivatives, since they're operating on functions of a single variable (x or t). This may not seem

particularly helpful, but look at what happens if you divide each term in this equation by $\psi(x)T(t)$:

$$\frac{1}{\cancel{\psi(x)}\cancel{T(t)}} i\hbar \cancel{\psi(x)} \frac{d[T(t)]}{dt} = -\frac{1}{\psi(x)\cancel{T(t)}} \frac{\hbar^2 \cancel{T(t)}}{2m} \frac{d^2[\psi(x)]}{dx^2} + \frac{1}{\cancel{\psi(x)}\cancel{T(t)}} V[\cancel{\psi(x)}\cancel{T(t)}]$$

$$i\hbar \frac{1}{T(t)} \frac{d[T(t)]}{dt} = -\frac{\hbar^2}{2m} \frac{1}{\psi(x)} \frac{d^2[\psi(x)]}{dx^2} + V. \tag{3.36}$$

Now consider each side of this equation in any case for which the potential energy (V) depends only on location and not on time. In such cases, the left side of this equation is a function only of time, and the right side is a function only of location. For that to be true, each side must be *constant*.

To see why that's the case, imagine what would happen at a given location (that is, a fixed value of x) if the left side of this equation changed over time. In that case, the right side of the equation would not be changing (since it depends only on location, and the location isn't changing), while the left side *would* be changing, since it depends on t. Likewise, if the right side of the equation changed over distance, at a fixed time (t not changing), moving to a different location would cause the right side of this equation to vary while the left side would not change. So for this equation to hold true, *both* sides must be constant.

Many students find this a bit troubling – isn't the wavefunction $\Psi(x,t)$ a function of both location (x) and time (t)? Yes, it is. But remember that we're not saying that the wavefunction $\Psi(x,t)$ and its derivatives don't change over space and time, it's $i\hbar \frac{1}{T(t)} \frac{d[T(t)]}{dt}$ (the left side) and $-\frac{\hbar^2}{2m} \frac{1}{\psi(x)} \frac{d^2[\psi(x)]}{dx^2} + V$ (the right side) that must be constant. And that's very different from $\Psi(x,t)$ or its derivatives being constant.

So what does it mean if each side of the equation is constant? Look first at the left side:

$$i\hbar \frac{1}{T(t)} \frac{d[T(t)]}{dt} = (\text{constant})$$

$$\frac{1}{T(t)} \frac{d[T(t)]}{dt} = \frac{(\text{constant})}{i\hbar}. \tag{3.37}$$

Integrating both sides of this equation over time gives

$$\int_0^t \frac{1}{T(t)} \frac{d[T(t)]}{dt} dt = \int_0^t \frac{(\text{constant})}{i\hbar} dt$$

$$\ln[T(t)] = \frac{(\text{constant})}{i\hbar} t = \frac{-\text{i}(\text{constant})}{\hbar} t$$

$$T(t) = e^{-i\frac{(\text{constant})}{\hbar}t}.$$

Calling the constant E (the reason for this choice will be explained later in the section)[6] makes this

$$T(t) = e^{-i\frac{E}{\hbar}t}. \tag{3.38}$$

So this is the solution to the time-function $T(t)$ portion of $\Psi(x,t)$, which tells you how the wavefunction evolves over time. You'll see this again after we've looked at the spatial function $\psi(x)$, for which the equation is

$$-\frac{\hbar^2}{2m}\frac{1}{\psi(x)}\frac{d^2[\psi(x)]}{dx^2} + V = E. \tag{3.39}$$

In this equation, the separation constant (E) must be the same as in the time equation (Eq. 3.37), since the two sides of Eq. 3.36 are equal. Multiplying all terms in Eq. 3.39 by $\psi(x)$ gives

$$-\frac{\hbar^2}{2m}\frac{d^2[\psi(x)]}{dx^2} + V[\psi(x)] = E[\psi(x)]. \tag{3.40}$$

This equation is called the time-independent Schrödinger equation (TISE), since its solutions $\psi(x)$ describe only the spatial behavior of the quantum wavefunction $\Psi(x,t)$ (the temporal behavior is described by the T(t) functions). And although the solutions to the TISE depend on the nature of the potential energy (V) in the region of interest, you can learn a lot just by looking carefully at Eq. 3.40.

The first thing to notice is that this is an eigenvalue equation. To see that, consider the operator

$$\widehat{H} = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + V. \tag{3.41}$$

As mentioned in Section 3.1, this is the one-dimensional version of the Hamiltonian (total energy) operator. Using this in the TISE makes it look like this:

$$\widehat{H}[\psi(x)] = E[\psi(x)], \tag{3.42}$$

which is exactly the form you'd expect for an eigenvalue equation, with eigenfunction $\psi(x)$ and eigenvalue E. This is why many authors refer to the process of solving the TISE as "finding the eigenvalues and eigenfunctions of the Hamiltonian operator."

[6] Even at this early stage of the discussion of the time function, you can tell that the constant must have dimensions of energy, since $-i\frac{\text{constant}}{\hbar}t$ must have dimensions of angle (radians), i is dimensionless, $\hbar$ has dimensions of (energy × time)/angle, and t has dimensions of time.

You're also likely to encounter the terminology "stationary states" for the functions $\psi(x)$, but it's important to realize that this does not mean that the wavefunctions $\Psi(x,t)$ are somehow "stationary" or not changing over time. Instead, this means that for any wavefunction $\Psi(x,t)$ that may be separated into spatial and temporal functions (as we did when we wrote $\Psi(x,t) = \psi(x)T(t)$ in Eq. 3.35), quantities such as the probability density and expectation values do not vary over time. To see why that's true, observe what happens when you form the inner product of such a separable wavefunction with itself:

$$\begin{aligned}\langle\Psi(x,t)|\Psi(x,t)\rangle \propto \Psi^*\Psi &= [\psi(x)T(t)]^*[\psi(x)T(t)] \\ &= [\psi(x)e^{-i\frac{E}{\hbar}t}]^*[\psi(x)e^{-i\frac{E}{\hbar}t}] \\ &= [\psi(x)]^*e^{i\frac{E}{\hbar}t}[\psi(x)]e^{-i\frac{E}{\hbar}t} = [\psi(x)]^*[\psi(x)],\end{aligned}$$

in which the time dependence has disappeared. Hence any quantity involving $\Psi^*\Psi$ will not change over time (it will become "stationary") whenever $\Psi(x,t)$ is separable.

You should also note that since the TISE is an eigenvalue equation, you can use the formalism of Chapter 2 to find and understand the meaning of the solutions to this equation. You can see how that works in Chapters 4 and 5, but before getting to that you may want to take a look at the three-dimensional version of the Schrödinger equation, which is the subject of the final section of this chapter.

3.4 Three-Dimensional Schrödinger Equation

Up to this point, we've been taking the spatial variation of the quantum wavefunction to depend on the single variable x, but many interesting problems in quantum mechanics are three-dimensional in nature. As you have probably surmised, extending the Schrödinger equation to three dimensions involves writing the wavefunction as $\Psi(\vec{r},t)$ rather than $\Psi(x,t)$.

This change is necessary because, in the one-dimensional case, position can be specified by the scalar x, but to specify position in three dimensions requires a position vector with three components, each pertaining to a different basis vector. For example, in 3-D Cartesian coordinates, the position vector $\vec{r}$ can be expressed using the orthonormal basis vectors ($\hat{\imath}$, $\hat{\jmath}$, and $\hat{k}$):

$$\vec{r} = x\hat{\imath} + y\hat{\jmath} + z\hat{k}. \tag{3.43}$$

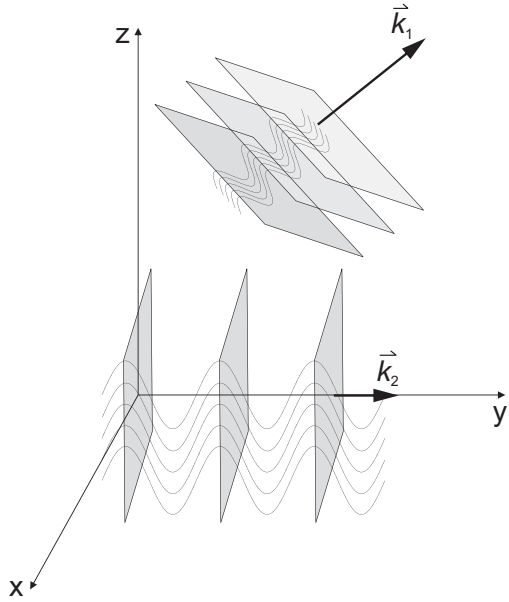


Figure 3.4 Three-dimensional plane waves.

Likewise, in the 1-D case the direction of propagation of the wave is constrained to a single axis, which meant we could use the scalar wavenumber k. But in the 3-D case, the wave may propagate in any direction, as shown in Fig. 3.4. That means that the wavenumber becomes a vector $\vec{k}$, which can be expressed using vector components k_x, k_y, and k_z as

$$\vec{k} = k_x\hat{\imath} + k_y\hat{\jmath} + k_z\hat{k} \tag{3.44}$$

in the 3-D Cartesian coordinate system. Note that the relationship between the magnitude of the vector wavenumber ($|\vec{k}|$) and the wavelength (λ) is preserved:

$$|\vec{k}| = \sqrt{k_x^2 + k_y^2 + k_z^2} = \frac{2\pi}{\lambda}. \tag{3.45}$$

Introducing the 3-D position vector $\vec{r}$ and propagation vector $\vec{k}$ into the plane-wave function Ψ results in an expression like this:

$$\Psi(\vec{r}, t) = Ae^{i(\vec{k}\circ\vec{r}-\omega t)}, \tag{3.46}$$

in which $\vec{k} \circ \vec{r}$ represents the scalar product between vectors $\vec{r}$ and $\vec{k}$.

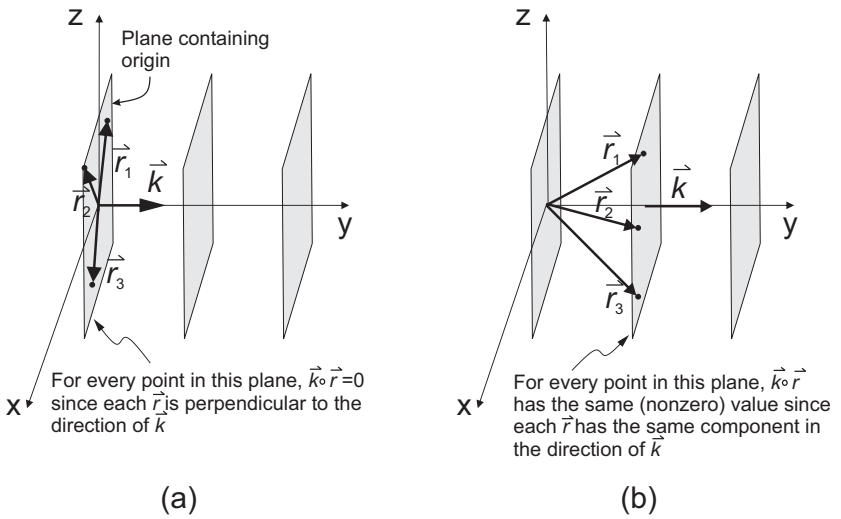


Figure 3.5 Plane-wave dot product for points in (a) plane containing origin and (b) plane displaced from origin.

If you're wondering why a dot product appears in this expression, take a look at the illustration of a plane wave propagating along the y-axis in Fig. 3.5.

As shown in this figure, the surfaces of constant phase are planes that are perpendicular to the direction of propagation, so these planes are parallel to the xz-plane in this case. For clarity, only those planes at the positive peak of the sinusoidal function are shown, but you can imagine similar planes existing at any other phase (or all other phases) of the wave.

The relevant point is that over each of these planes, the dot product $\vec{k} \circ \vec{r}$ gives the same numerical value for every position vector between the origin and any point in the plane. That's probably easiest to see in the plane that passes through the origin, as shown in Fig. 3.5a. Since the position vectors for all points in that plane are perpendicular to the propagation vector $\vec{k}$, the dot product $\vec{k} \circ \vec{r}$ has the constant value of zero for that plane.

Now consider the position vectors from the origin to points in the next plane to the right, as shown in Fig. 3.5b. Remember that the dot product between two vectors gives a result that's proportional to the projection of one of the vectors onto the direction of the other. Since each of the position vectors to points in this plane have the same y-component, the dot product $\vec{k} \circ \vec{r}$ has a constant nonzero value for this plane.

What exactly is that value? Note that $\vec{k} \circ \vec{r} = |\vec{k}||\vec{r}|cos(\theta)$, in which θ is the angle between vectors $\vec{k}$ and $\vec{r}$. Note also that $|\vec{r}|cos(\theta)$ is the distance from the origin to the point on the plane closest to the origin along the direction of $\vec{k}$ (that is, the perpendicular distance from the origin to the plane). So this dot product gives the distance from the origin to the plane along the direction of $\vec{k}$ multiplied by $|\vec{k}|$. And since $|\vec{k}| = \frac{2\pi}{\lambda}$, multiplying any distance by the magnitude of k has the effect of dividing that distance by the wavelength λ (which tells you how many wavelengths fit into that distance) and then multiplying that result by 2π (which converts the number of wavelengths into radians, since each wavelength represents 2π radians of phase).

Extending the same logic to any other plane of constant phase provides the reason for the appearance of the dot product $\vec{k} \circ \vec{r}$ in the 3-D wavefunction $\Psi(\vec{r}, t)$: it gives the distance from the origin to the plane in units of radians, which is exactly what's needed to account for the variation in the phase of $\Psi(\vec{r}, t)$ as the wave propagates in the $\vec{k}$ direction.

So that's why $\vec{k} \circ \vec{r}$ appears in the 3-D wavefunction, and you may see that dot product expanded in Cartesian coordinates as

$$\begin{aligned}\vec{k} \circ \vec{r} &= (k_x\hat{\imath} + k_y\hat{\jmath} + k_z\hat{k}) \circ (x\hat{\imath} + y\hat{\jmath} + z\hat{k}) \\ &= k_x x + k_y y + k_z z,\end{aligned}$$

which makes the 3-D plane-wave function in Cartesian coordinates look like this:

$$\Psi(\vec{r}, t) = Ae^{i[(k_x x + k_y y + k_z z) - \omega t]}. \tag{3.47}$$

In addition to extending the wavefunction $\Psi(x, t)$ to three dimensions as $\Psi(\vec{r}, t)$, it's also necessary to extend the second-order spatial derivative $\frac{\partial^2}{\partial x^2}$ to three dimensions. To see how to do that, start by taking the first and second spatial derivatives of $\Psi(\vec{r}, t)$ with respect to x:

$$\begin{aligned}\frac{\partial \Psi(\vec{r}, t)}{\partial x} &= \frac{\partial \left[Ae^{i[(k_x x + k_y y + k_z z) - \omega t]}\right]}{\partial x} = ik_x \left[Ae^{i[(k_x x + k_y y + k_z z) - \omega t]}\right] \\ &= ik_x \Psi(\vec{r}, t)\end{aligned}$$

and

$$\begin{aligned}\frac{\partial^2 \Psi(\vec{r}, t)}{\partial x^2} &= \frac{\partial \left[ik_x Ae^{i[(k_x x + k_y y + k_z z) - \omega t]}\right]}{\partial x} = ik_x \left[ik_x Ae^{i[(k_x x + k_y y + k_z z) - \omega t]}\right] \\ &= -k_x^2 \Psi(\vec{r}, t).\end{aligned}$$

The second spatial derivatives with respect to y and z are

$$\frac{\partial^2 \Psi(\vec{r},t)}{\partial y^2} = -k_y^2 \Psi(\vec{r},t)$$

and

$$\frac{\partial^2 \Psi(\vec{r},t)}{\partial z^2} = -k_z^2 \Psi(\vec{r},t).$$

Adding these second derivatives together gives

$$\begin{aligned}\frac{\partial^2 \Psi(\vec{r},t)}{\partial x^2} + \frac{\partial^2 \Psi(\vec{r},t)}{\partial y^2} + \frac{\partial^2 \Psi(\vec{r},t)}{\partial z^2} &= -k_x^2 \Psi(\vec{r},t) - k_y^2 \Psi(\vec{r},t) - k_z^2 \Psi(\vec{r},t) \\ &= -(k_x^2 + k_y^2 + k_z^2)\Psi(\vec{r},t),\end{aligned}$$

and you know from Eq. 3.45 that the sum of the squares of the components of $\vec{k}$ gives the square of the magnitude of $\vec{k}$, so

$$\frac{\partial^2 \Psi(\vec{r},t)}{\partial x^2} + \frac{\partial^2 \Psi(\vec{r},t)}{\partial y^2} + \frac{\partial^2 \Psi(\vec{r},t)}{\partial z^2} = -|\vec{k}|^2 \Psi(\vec{r},t). \tag{3.48}$$

Comparing this equation to Eq. 3.16 shows that the sum of the second spatial derivatives brings down a factor of $-|\vec{k}|^2$ from the plane-wave exponential, just as $\frac{\partial^2 \Psi(x,t)}{\partial x^2}$ brought down a factor of $-k^2$ in the one-dimensional case.

This sum of second spatial derivatives can be written as a differential operator:

$$\frac{\partial^2 \Psi(\vec{r},t)}{\partial x^2} + \frac{\partial^2 \Psi(\vec{r},t)}{\partial y^2} + \frac{\partial^2 \Psi(\vec{r},t)}{\partial z^2} = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \Psi(\vec{r},t).$$

This is the Cartesian version of the **Laplacian operator** (sometimes called the "del-squared" operator), which most texts write using this notation[7]:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}. \tag{3.49}$$

With the Laplacian operator ∇^2 and the three-dimensional wavefunction $\Psi(\vec{r},t)$ in hand, you can write the Schrödinger equation as

$$i\hbar \frac{\partial \Psi(\vec{r},t)}{\partial t} = -\frac{\hbar^2}{2m} \nabla^2 \Psi(\vec{r},t) + V[\Psi(\vec{r},t)]. \tag{3.50}$$

This three-dimensional version of the time-dependent Schrödinger equation shares several features with the one-dimensional version, but there are a few

[7] In some texts you'll see the Laplacian written as Δ instead of ∇^2.

subtleties in the interpretation of the Laplacian that bear further examination. As in the one-dimensional case, comparison to the diffusion equation is a good place to begin. The three-dimensional version of the diffusion equation is

$$\frac{\partial[f(\vec{r},t)]}{\partial t} = D\nabla^2[f(\vec{r},t)]. \tag{3.51}$$

Just as in the one-dimensional case, this three-dimensional diffusion equation describes the behavior of a quantity $f(\vec{r},t)$ with spatial distribution that may evolve over time, and again the proportionality factor "D" between the first-order time derivative $\frac{\partial f}{\partial t}$ and the second-order spatial derivatives $\nabla^2 f$ represents the diffusion coefficient.

To see the similarity between the 3-D diffusion equation and the 3-D Schrödinger equation, consider again the case in which the potential energy (V) is zero, and write Eq. 3.50 as

$$\frac{\partial[\Psi(\vec{r},t)]}{\partial t} = \frac{i\hbar}{2m}\nabla^2[\Psi(\vec{r},t)]. \tag{3.52}$$

As in the 1-D case, the presence of the "i" factor in the Schrödinger equation has important implications, but the fundamental relationship in both of these equations is this: the evolution of the wavefunction over time is proportional to the Laplacian of the wavefunction.

To understand the nature of the Laplacian operator, it helps to view spatial curvature from another perspective. That perspective is to consider how the value of a function at a given point compares to the average value of that function at equidistant neighboring points.

This concept is straightforward for a one-dimensional function $\psi(x)$, which could represent, for example, the temperature distribution along a bar. As you can see in Fig. 3.6, the curvature of the function determines whether the value of the function at any point is equal to, greater than, or less than the average value of the function at equidistant surrounding points.

Consider first the zero-curvature case shown in Fig. 3.6a. Zero curvature means that the slope of $\psi(x)$ is constant in this region, so the value of ψ at position x_0 lies on a straight line between the values of ψ at equal distances on opposite sides of position x_0. That means that the value of $\psi(x_0)$ must be equal to the average of the values of ψ at positions an equal distance (shown as Δx in the figure) on either side of x_0. So in this case $\psi(x_0) = \frac{1}{2}[\psi(x_0 + \Delta x) + \psi(x_0 - \Delta x)]$.

But if the function $\psi(x)$ has positive curvature as shown in Fig. 3.6b, the value of $\psi(x)$ at position x_0 is less than the average of the values of the function at equidistant positions $x_0 + \Delta x$ and $x_0 - \Delta x$. Hence for positive curvature

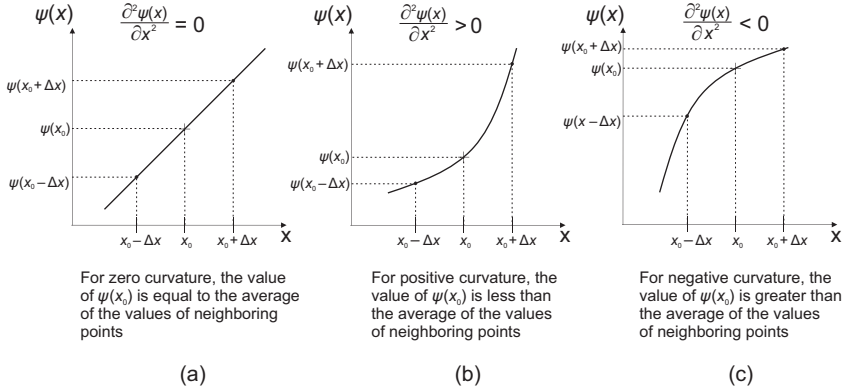


Figure 3.6 Laplacian for (a) zero, (b) positive, and (c) negative curvature.

$\psi(x_0) < \frac{1}{2}[\psi(x_0 + \Delta x) + \psi(x_0 - \Delta x)]$, and the more positive the curvature, the greater the amount by which $\psi(x_0)$ falls short of the average of surrounding points.

Likewise, if the function $\psi(x)$ has negative curvature as shown in Fig. 3.6c, the value of $\psi(x)$ at position x_0 is greater than the average of the values of the function at equidistant positions $x_0 + \Delta x$ and $x_0 - \Delta x$. So for negative curvature $\psi(x_0) > \frac{1}{2}[\psi(x_0 + \Delta x) + \psi(x_0 - \Delta x)]$, and the more negative the curvature, the greater the amount by which $\psi(x_0)$ exceeds the average value of $\psi(x)$ at surrounding points.

The bottom line is that the curvature of a function at any location is a measure of the amount by which the value of the function at that location equals, exceeds, or falls short of the average value of the function at surrounding points.

To extend this logic to functions of more than one spatial dimension, consider the two-dimensional function $\psi(x, y)$. This function might represent the temperature at various points (x, y) on a slab, the concentration of particulates on the surface of a stream, or the height of the ground above some reference surface such as sea level.

Two-dimensional functions can be conveniently plotted in three dimensions, as shown in Fig. 3.7. In this type of plot, the z-axis represents the quantity of interest, such as temperature, concentration, or height above sea level in the examples just mentioned.

Consider first the function shown in Fig. 3.7a, which has a positive peak (maximum value) at position $(x = 0, y = 0)$. The value $\psi(0, 0)$ of this function

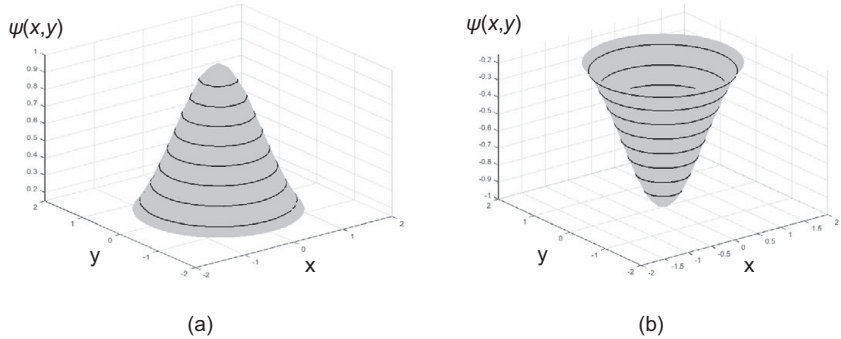


Figure 3.7 Two-dimensional function $\psi(x, y)$ with contours for (a) maximum at origin and (b) minimum at origin.

at the peak definitely exceeds the average value of the function at equidistant surrounding points, such as the points along the circular contours shown in the figure. That's consistent with the negative-curvature case in the 1-D example discussed earlier in the section.

Now look at the function shown in Fig. 3.7b, which has a circular valley (minimum value) at position ($x = 0$, $y = 0$). In this case, the value $\psi(0, 0)$ of this function at the center of the valley definitely falls short of the average value of the function at equidistant surrounding points, consistent with the positive-curvature case in the 1-D example.

By imagining cuts through the function $\psi(x, y)$ along the x- and y-directions, you may be able to convince yourself that near the peak of the positively peaked function in Fig. 3.7, the curvature is negative, since the slope decreases as you move along each axis (that is, $\partial/\partial x(\frac{\partial \psi}{\partial x})$ and $\partial/\partial y(\frac{\partial \psi}{\partial y})$ are both negative).

But another way of understanding the behavior of 2-D functions is to consider the Laplacian as the combination of two differential operators: the gradient and the divergence. You may have encountered these operators in a multivariable calculus or electromagnetics class, but don't worry if you're not clear on their meaning – the following explanation should help you understand them and their role in the Laplacian.

In colloquial speech, the word "gradient" is typically used to describe the change in some quantity with position, such as the change in the height of a sloping road, the variation in the intensity of a color in a photograph, or the increase or decrease in temperature at different locations in a room. Happily,

that common usage provides a good basis for the mathematical definition of the gradient operator, which looks like this in 3-D Cartesian coordinates:

$$\vec{\nabla} = \hat{\imath}\frac{\partial}{\partial x} + \hat{\jmath}\frac{\partial}{\partial y} + \hat{k}\frac{\partial}{\partial z}, \tag{3.53}$$

in which the symbol ∇ is called "del" or "nabla." In case you're wondering, the reason for writing the unit vectors ($\hat{\imath}$, $\hat{\jmath}$, and $\hat{k}$) to the left of the partial derivatives is to make it clear that those derivatives are meant to operate on whatever function you feed the operator; the derivatives are not meant to operate on the unit vectors.

As with any operator, the del operator doesn't do anything until you feed it something on which it can operate. So the gradient of function $\psi(x, y, z)$ in Cartesian coordinates is

$$\vec{\nabla}\psi(x, y, z) = \frac{\partial \psi}{\partial x}\hat{\imath} + \frac{\partial \psi}{\partial y}\hat{\jmath} + \frac{\partial \psi}{\partial z}\hat{k}. \tag{3.54}$$

From this definition, you can see that taking the gradient of a scalar function (such as ψ) produces a vector result, and both the direction and the magnitude of that vector are meaningful. The direction of the gradient vector tells you the direction of steepest increase in the function, and the magnitude of the gradient tells you the rate of change of the function in that direction.

You can see the gradient in action in Fig. 3.8. Since the gradient vectors point in the direction of steepest increase of the function, they point "uphill" toward the peak in the (a) portion of the figure and away from the bottom of the valley in the (b) portion of the figure. And since contours represent lines

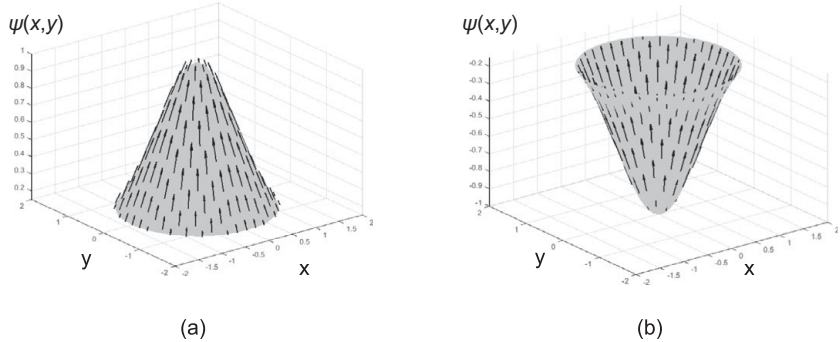


Figure 3.8 Gradients for (a) 2-D peak and (b) 2-D valley functions.

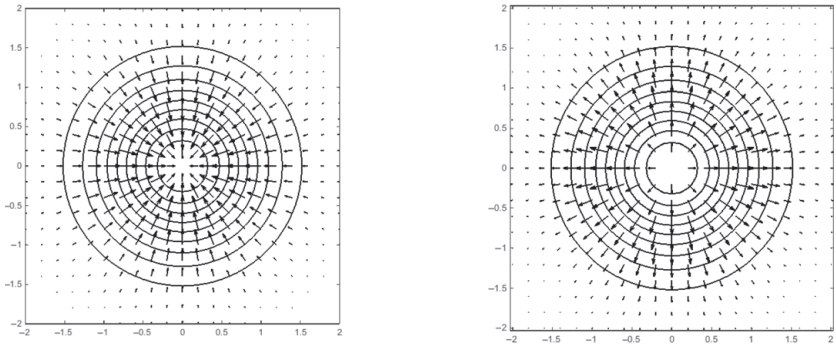


Figure 3.9 Top view of contours and gradients for 2-D (a) peak and (b) valley functions.

of constant value of the function ψ, the direction of the gradient vectors must always be perpendicular to the contours (those contours are shown in Fig. 3.7).

To understand the role of the gradient in the Laplacian, it may help you to consider the top views of the gradients of the peak and valley functions, which are shown in Fig. 3.9. From this viewpoint, you can see that the gradient vectors converge toward the top of the positive peak and diverge away from the bottom of the valley (and are perpendicular to equal-value contours, as mentioned in the previous paragraph).

The reason this top view is useful is that it makes clear the role of another operator that works in tandem with the gradient to produce the Laplacian. That operator is the divergence, which is written as a scalar (dot) product between the gradient operator $\vec{\nabla}$ and a vector (such as $\vec{A}$). In 3-D Cartesian coordinates, that means

$$\begin{aligned}\vec{\nabla} \circ \vec{A} &= \left(\hat{\imath}\frac{\partial}{\partial x} + \hat{\jmath}\frac{\partial}{\partial y} + \hat{k}\frac{\partial}{\partial z}\right) \circ (A_{\mathrm{x}}\hat{\imath} + A_{\mathrm{y}}\hat{\jmath} + A_{\mathrm{z}}\hat{k}) \\ &= \frac{\partial A_{\mathrm{x}}}{\partial x} + \frac{\partial A_{\mathrm{y}}}{\partial y} + \frac{\partial A_{\mathrm{z}}}{\partial z}. \end{aligned} \tag{3.55}$$

Note that the divergence operates on a vector function and produces a scalar result.

And what does the scalar result of taking the divergence of a vector function tell you about that function? At any location, the divergence tells you whether the function is diverging (loosely meaning "spreading out") or converging

(loosely meaning "coming together") at that point. One way to visualize the meaning of the divergence of a vector function is to imagine that the vectors represent the velocity vectors of a flowing fluid. At a location of large positive divergence, more fluid flows away from that location than toward it, so the flow vectors diverge from that location (and a "source" of fluid exists at that point). For locations with zero divergence, the fluid flow away from that location exactly equals the fluid flow toward it. And as you might expect, locations with large negative divergence have more fluid flowing toward them than away from them (and a "sink" of fluid exists at that point).

Of course, most vector fields don't represent the flow of a fluid, but the concept of vector "flow" toward or away from a point is still useful. Just imagine a tiny sphere surrounding the point of interest, and determine whether the outward flux of the vector field (which you can think of as the number of vectors that cross the surface from inside to outside) is greater than, equal to, or less than the inward flux (the number of vectors that cross the surface from outside to inside).

One oft-cited thought experiment to test the divergence at a given point using the fluid-flow analogy is to imagine sprinkling loose material such as sawdust or powder into the flowing fluid. If the sprinkled material disperses (that is, if its density decreases), then the divergence at that location is positive. But if the sprinkled material compresses (that is, its density increases), then the divergence at that location is negative. And if the material neither disperses nor compresses but simply retains its original density as it moves along with the flow, then the divergence is zero at that location.

It may seem that we've wandered quite far from the Laplacian and the diffusion equation, but here's the payoff: The Laplacian of a function ($\nabla^2\psi$) is identical to the divergence of the gradient of that function ($\vec{\nabla} \circ \vec{\nabla}\psi$). You can see that by taking the dot product of the divergence and the gradient of ψ:

$$\begin{aligned}\vec{\nabla} \circ \vec{\nabla}\psi &= \frac{\partial\left(\frac{\partial\psi}{\partial x}\right)}{\partial x} + \frac{\partial\left(\frac{\partial\psi}{\partial y}\right)}{\partial y} + \frac{\partial\left(\frac{\partial\psi}{\partial z}\right)}{\partial z} \\ &= \frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2} = \nabla^2\psi.\end{aligned}$$

So the divergence of the gradient is equivalent to the Laplacian. This ties together the interpretation of gradient vectors converging on a peak (which means that the divergence of the gradient is negative at a peak) with the value at the peak being greater than the average value of the surrounding points (which means that the Laplacian is negative at a peak).

And what does all this have to do with the diffusion equation and the Schrödinger equation? Recall that the diffusion equation states that the change over time of function ψ (that is, $\frac{\partial \psi}{\partial t}$) is proportional to the Laplacian of ψ (given by $\nabla^2 \psi$). So if ψ represents temperature, diffusion will cause any region in which the temperature exceeds the average temperature at surrounding points (that is, a region in which the function ψ has a positive peak) to cool down, while any region in which the temperature is lower than the average temperature at surrounding points (where function ψ has a valley) will warm up.

A similar analysis applies to the Schrödinger equation, with one very important difference. Just as in the one-dimensional case, the presence of the imaginary unit ("i") on one side of the Schrödinger equation means that the solutions will generally be complex rather than purely real. That means that in addition to "diffusing" solutions in which the peaks and valleys of the function tend to smooth out over time, oscillatory solutions are also supported. You can read about those solutions in Chapters 4 and 5.

Before getting to that, it's worth noting that a three-dimensional version of the time-independent Schrödinger equation (TISE) can be found using an approach similar to that used in the one-dimensional case. To see that, separate the 3-D wavefunction $\Psi(\vec{r}, t)$ into spatial and temporal parts:

$$\Psi(\vec{r}, t) = \psi(\vec{r})T(t), \tag{3.56}$$

and write a 3-D version of the potential energy $V(\vec{r})$. Just as in the 1-D case, the time portion of the equation leads to the solution $T(t) = e^{-i\frac{E}{\hbar}t}$, but the 3-D spatial-portion equation is

$$-\frac{\hbar^2}{2m}\nabla^2[\psi(\vec{r})] + V[\psi(\vec{r})] = E[\psi(\vec{r})]. \tag{3.57}$$

The solutions to this 3-D TISE depend on the nature of the potential $V(\vec{r})$, and in this case the 3-D version of the Hamiltonian (total energy) operator is

$$\widehat{H} = -\frac{\hbar^2}{2m}\nabla^2 + V. \tag{3.58}$$

One final note on the Laplacian operator that appears in the 3-D Schrödinger equation: although the Cartesian version of the Laplacian has the simplest form, the geometry of some problems (specifically those with spherical symmetry) suggests that the Laplacian operator written in spherical coordinates may be easier to apply. That version looks like this:

$$\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\phi^2}, \qquad (3.59)$$

and you can see an application of this version of the Laplacian in the chapter-end problems and online solutions.

3.5 Problems

1. Find the deBroglie wavelength of the matter wave associated with

a) An electron travelling at a speed of 5×10^6 m/s.

b) A 160-gram cricket ball bowled at a speed of 100 miles per hour.

2. Given the wavenumber function $\phi(k) = A$ for the wavenumber range $-\frac{\Delta k}{2} < k < \frac{\Delta k}{2}$ and zero elsewhere, use Eq. 3.24 to find the corresponding position wavefunction $\psi(x)$. A represents a constant.

3. Find the matrix representation of the momentum operator $\widehat{p}$ in a 2-D basis system with basis vectors represented by kets $|\epsilon_1\rangle = \sin kx$ and $|\epsilon_2\rangle = \cos kx$.

4. Find the matrix representation of the Hamiltonian operator $\widehat{H}$ in a region with constant potential energy V for the same 2-D basis system as Problem 3.

5. Show that the momentum operator and Hamiltonian operator with constant potential energy commute using the functional representations (Eqs. 3.29 and 3.30) and using the matrix representations of these operators in the basis system given in Problems 3 and 4.

6. For a plane wave with vector wavenumber $\vec{k} = \hat{\imath} + \hat{\jmath} + 5\hat{k}$,

a) Sketch a few of the planes of constant phase for this wave using 3-D Cartesian coordinates.

b) Find the wavelength λ of this wave.

c) Determine the minimum distance from the origin to the plane containing the point $(x = 4, y = 2, z = 5)$ along the direction of $\vec{k}$.

7. For the 2-D Gaussian wavefunction $f(x, y) = Ae^{-\left[\frac{(x-x_0)^2}{2\sigma_x^2} + \frac{(y-y_0)^2}{2\sigma_y^2}\right]}$, show that

a) The gradient $\vec{\nabla} f$ is zero at the peak of the function $(x = x_0, y = y_0)$.

b) The Laplacian $\nabla^2 f$ is negative at the location of the peak.

c) The sharper the peak (smaller σ_x and σ_y), the larger the Laplacian.

8. Show that $\Psi_n(x, y, z, t) = \sqrt{\frac{8}{a_x a_y a_z}} \sin\left(k_{n,x}x\right) \sin\left(k_{n,y}y\right) \sin\left(k_{n,z}z\right) e^{-iE_n t/\hbar}$ is a solution to the Schrödinger equation in 3-D Cartesian coordinates if $E_n = \frac{k_n^2 \hbar^2}{2m}$ with $k_n^2 = (k_{n,x})^2 + (k_{n,y})^2 + (k_{n,z})^2$ and $k_{n,x} = \frac{n_x \pi}{a_x}$, $k_{n,y} = \frac{n_y \pi}{a_y}$, and $k_{n,z} = \frac{n_z \pi}{a_z}$ in a region of constant potential.

9. Use separation of variables to write the 3-D Schrödinger equation in spherical coordinates as two separate equations, one depending only on the radial coordinate (r) and the other depending only on the angular coordinates (θ and ϕ), with the potential energy depending only on the radial coordinate (so $V = V(r)$).

10. Show that the function $R(r) = \frac{1}{r\sqrt{2\pi a}} \sin\left(\frac{n\pi r}{a}\right)$ is a solution to the radial portion of the 3-D Schrödinger equation in spherical coordinates for $V = 0$ and with separation constant $E_n = \frac{n^2 \pi^2 \hbar^2}{2ma^2}$.

4

Solving the Schrödinger Equation

If you're wondering how the abstract vector spaces, orthogonal functions, operators, and eigenvalues discussed in Chapters 1 and 2 relate to the wavefunction solutions to the Schrödinger equation developed in Chapter 3, you should find this chapter helpful. One reason that relationship may not be obvious is that quantum mechanics was developed along two parallel paths, which have come to be called the "**matrix mechanics**" of Werner Heisenberg and the "**wave mechanics**" of Erwin Schrödinger. And although those two approaches are known to yield equivalent results, each offers benefits in elucidating certain aspects of quantum theory. That's why Chapters 1 and 2 focused on matrix algebra and Dirac notation while Chapter 3 dealt with plane waves and differential operators.

To help you understand the connections between matrix mechanics and wave mechanics, the first section of this chapter explains the meaning of the solutions to the Schrödinger equation using the Born rule, which is the basis for the Copenhagen interpretation of quantum mechanics. In Section 4.2, you'll find a discussion of quantum states, wavefunctions, and operators, along with an explanation of several dangerous misconceptions that are commonly held by students attempting to apply quantum theory to practical problems.

The requirements and general characteristics of quantum wavefunctions are discussed in Section 4.3, after which you can see how Fourier theory applies to quantum wavefunctions in Section 4.4. The final section of this chapter presents and explains the form of the position and momentum operators in both position and momentum space.

4.1 The Born Rule and Copenhagen Interpretation

When Schrödinger published his equation in early 1926, no one (including Schrödinger himself) knew with certainty what the wavefunction ψ represented. Schrödinger thought that the wavefunction of a charged particle might be related to the spatial distribution of electric charge density, suggesting a literal interpretation of the wavefunction as a real disturbance – a "matter wave." Others speculated that the wavefunction might represent some type of "guiding wave" that accompanies every physical particle and controls certain aspects of its behavior. Each of these ideas has some merit, but the question of what is actually "waving" in the quantum wavefunction solutions to the Schrödinger equation was very much open to debate.

The answer to that question came later in 1926, when Max Born published a paper in which he stated what he believed was the only possible interpretation of the wavefunction solution to the Schrödinger equation. That answer, now known as the "Born rule," says that the quantum wavefunction represents a "probability amplitude" whose magnitude squared determines the probability that a certain result will be obtained when an observation is made. You can read more about wavefunctions and probability in the next section, but for now the important point is that the Born rule removes the quantum wavefunction from the realm of measurable disturbances in a physical medium, and relegates ψ to the world of statistical tools (albeit very useful ones). Specifically, the wavefunction may be used to determine the possible results of measurements of quantum observables and to calculate the probabilities of each of those results.

The Born rule plays an extremely important role in quantum mechanics, since it explains the meaning of the solutions to the Schrödinger equation in a way that matches experimental results. But the Born rule is silent about other critical aspects of quantum mechanics, and unlike the almost immediate and widespread acceptance of the Born rule, those other aspects have been the subject of continuing debate for nearly a century.

That debate has not led to a set of universally agreed-upon principles. The most widely accepted (and widely disputed) explanation of quantum mechanics is called the "Copenhagen interpretation," since it was developed in large part at the Niels Bohr Institute in Copenhagen. In spite of the ambivalence many quantum theorists express toward the Copenhagen interpretation, it's worth your time to understand its basic tenets. With that understanding, you'll be able to appreciate the features and drawbacks of the Copenhagen interpretation, as well as the advantages and difficulties of alternative interpretations.

So exactly what are those tenets? That's not easy to say, since there seem to be almost as many versions of the Copenhagen interpretation as there are bicycles in Copenhagen. But the principles usually attributed to the Copenhagen interpretation include the completeness of the information in the quantum state, the smooth time evolution of quantum states, wavefunction collapse, the relationship of operator eigenvalues to measurement results, the uncertainty principle, the Born rule, the correspondence principle between classical and quantum physics, and the complementary wave and particle aspects of matter.

Here's a short description of each of these principles:

Information content The quantum state Ψ includes all possible information about a quantum system – there are no "hidden variables" with additional information.

Time evolution Over time, quantum states evolve smoothly in accordance with the Schrödinger equation unless a measurement is made.

Wavefunction collapse Whenever an measurement of a quantum state is made, the state "**collapses**" to an eigenstate of the operator associated with the observable being measured.

Measurement results The value measured for an observable is the eigenvalue of the eigenstate to which the original quantum state has collapsed.

Uncertainty principle Certain "incompatible" observables (such as position and momentum) may not be simultaneously known with arbitrarily great precision.

Born rule The probability that a quantum state will collapse to a given eigenstate upon measurement is determined by the square of the amount of that eigenstate present in the original state (the wavefunction).

Correspondence principle In the limit of very large quantum numbers, the results of measurements of quantum observables must match the results of classical physics.

Complementarity Every quantum system includes complementary wave-like and particle-like aspects; whether the system behaves like a wave or like a particle when measured is determined by the nature of the measurement.

Happily, whether you favor the Copenhagen interpretation or one of the alternative explanations, the "mechanics" of quantum mechanics works. That is, the quantum-mechanical techniques for predicting the outcomes of measurements of quantum observables and calculating the probability of each of those outcomes has been repeatedly demonstrated to give correct answers.

You can read more about the quantum wavefunctions that are solutions of the Schrödinger equation later in this chapter, but in the next section you'll find a review of quantum terminology and a discussion of several common misconceptions about wavefunctions, operators, and measurements.

4.2 Quantum States, Wavefunctions, and Operators

As you may have observed in working through earlier chapters, some of the concepts and mathematical techniques of classical mechanics may be extended to the domain of quantum mechanics. But the fundamentally probabilistic nature of quantum mechanics leads to several profound differences, and it's very important for you to develop a firm grasp of those differences. That grasp includes an understanding of how certain classical-physics terminology does or does not apply to quantum mechanics.

Fortunately, progress has been made in developing consistent terminology in the roughly 100 years since the birth of quantum mechanics, but if you read the most popular quantum texts and online resources, you're likely to notice some variation in the use of the terms "quantum state" and "wavefunction." Although some authors use these terms interchangeably, others draw a significant distinction between them, and that distinction is explained in this section.

In the most common use of the term, the quantum state of a particle or system is a description that contains all the information that can be known about the particle or system. A quantum state is usually written as ψ (sometimes uppercase Ψ, especially when time dependence is included) and can be represented by a basis-independent ket $|\psi\rangle$ or $|\Psi\rangle$. Quantum states are members of an abstract vector space and obey the rules for such spaces, and the Schrödinger equation describes how a quantum state evolves over time.

So what's the difference between a quantum state and a quantum wavefunction? In a number of quantum texts, a quantum wavefunction is defined as the expansion of a quantum state in a specified basis. And which basis is that? Whichever basis you choose, and a logical choice is the basis corresponding to the observable of interest. Recall that every observable is associated with an operator, and the eigenfunctions of that operator form a complete orthogonal basis. That means that any function may be synthesized by weighted combination (superposition) of those eigenfunctions. As described in Section 1.6, if you expand the quantum state using a weighted sum of the eigenfunctions

$(\psi_1, \psi_2, \ldots, \psi_N)$ for that basis, then a state represented by ket $|\psi\rangle$ may be written as

$$|\psi\rangle = c_1 |\psi_1\rangle + c_2 |\psi_2\rangle + \cdots + c_N |\psi_N\rangle = \sum_{n=1}^{N} c_n |\psi_n\rangle , \tag{1.35}$$

and the wavefunction is the amount (c_n) of each eigenfunction $|\psi_n\rangle$ in state $|\psi\rangle$. So the wavefunction in a specified basis is the collection of (potentially complex) values c_n for that basis.

Also as described in Section 1.6, each c_n may be found by projecting state $|\psi\rangle$ onto the corresponding (normalized) eigenfunction $|\psi_n\rangle$:

$$c_n = \langle \psi_n | \psi \rangle . \tag{4.1}$$

The possible measurement outcomes are the eigenvalues of the operator corresponding to the observable, and the probability of each outcome is proportional to the square of the magnitude of the wavefunction value c_n. Thus the wavefunction represents the "probability amplitude" of each outcome.[1]

If it seems strange to apply the word "function" to a group of discrete values c_n, the reason for that terminology should become clear when you consider a quantum system (such as a **free particle**) in which the possible outcomes of measurements (the eigenvalues of the operator associated with the observable) are continuous functions rather than discrete values. In quantum textbooks, this is sometimes described as the operator having a "continuous spectrum" of eigenvalues.

In that case, the matrix representing the operator associated with an observable, such as position or momentum, has an infinite number of rows and columns, and there exist an infinite number of eigenfunctions for that observable. For example, the (one-dimensional) position basis functions may be represented by the ket $|x\rangle$, so expanding the state represented by ket $|\psi\rangle$ in the position basis looks like this:

$$|\psi\rangle = \int_{-\infty}^{\infty} \psi(x) |x\rangle \, dx. \tag{4.2}$$

Notice that the "amount" of the basis function $|x\rangle$ at each value of the continuous variable x is now the continuous function $\psi(x)$. So the wavefunction in this case is not a collection of discrete values (such as c_n), but rather the continuous function of position $\psi(x)$.

[1] The word "amplitude" is used in analogy with other types of waves, for which the intensity is proportional to the square of the wave's amplitude.

To determine $\psi(x)$, do exactly as you do in the discrete case: project the state $|\psi\rangle$ onto the position basis functions:

$$\psi(x) = \langle x|\psi\rangle\,. \tag{4.3}$$

Just as in the discrete case, the probability of each outcome is related to the square of the wavefunction. But in the continuous case $|\psi(x)|^2$ gives you the probability density (the probability per unit length in the 1-D case), which you must integrate over a range of x to determine the probability of an outcome within that range.

The same approach can be taken for the momentum wavefunction. The (one-dimensional) momentum basis functions may be represented by the ket $|p\rangle$, and expanding the state $|\psi\rangle$ in the momentum basis looks like this:

$$|\psi\rangle = \int_{-\infty}^{\infty} \tilde{\phi}(p)\,|p\rangle\,dp. \tag{4.4}$$

In this case the "amount" of the basis function at each value of the continuous variable p is the continuous function $\tilde{\phi}(p)$.

To determine $\tilde{\phi}(p)$, project the state $|\psi\rangle$ onto the momentum basis functions:

$$\tilde{\phi}(p) = \langle p|\psi\rangle\,. \tag{4.5}$$

So for a given quantum state represented by $|\psi\rangle$, the proper approach to finding the wavefunction in a specified basis is to use the inner product to project the quantum state onto the eigenfunctions for that basis. But research studies[2] have shown that even after completing an introductory course on quantum mechanics, many students are unclear on the relationship of quantum states, wavefunctions, and operators.

One common misconception concerning quantum operators is that if you're given a quantum state $|\psi\rangle$, you can determine the position-basis wavefunction $\psi(x)$ or the momentum-basis wavefunction $\tilde{\phi}(p)$ by operating on state $|\psi\rangle$ with the position or momentum operator. This is not true; as described previously, the correct way to determine the position or momentum waveform is to project the state $|\psi\rangle$ onto the eigenstates of position or momentum using the inner product.

A related misconception is that an operator may be used to convert between the position-basis wavefunction $\psi(x)$ and the momentum-basis wavefunction $\tilde{\phi}(p)$. But as you'll see in Section 4.4, the position-basis and momentum-basis

[2] See, for example, [4].

wavefunctions are related to one another by the **Fourier transform**, not by using the position or momentum operator.

It's also common for students new to quantum mechanics to believe that applying an operator to a quantum state is the analytical equivalent to making a physical measurement of the observable associated with that operator. Such confusion is understandable, since operating on a state does produce a new state, and many students have heard that making a measurement causes the collapse of a quantum wavefunction. The actual relationship between applying an operator and making a measurement is a bit more complex, but also more informative. The measurement of an observable does indeed cause a quantum state to collapse to one of the eigenstates of the operator associated with that observable (unless the state is already an eigenstate of that operator), but that is definitely not what happens when you apply an operator to a quantum state.

Instead, applying the operator produces a new quantum state that is the superposition (that is, the weighted combination) of the eigenstates of that operator. In that superposition of eigenstates, the weighting coefficient of each eigenstate is not just the "amount" (c_n) of that eigenstate, as it was in the expression for the state pre-operation:

$$|\psi\rangle = \Sigma_n c_n |\psi_n\rangle .$$

But after applying the operator to the quantum state, the weighting factor for each eigenstate includes the eigenvalue of that eigenstate, because the operator has brought out a factor of that eigenvalue (o_n):

$$\widehat{O}|\psi\rangle = \Sigma_n c_n \widehat{O}|\psi_n\rangle = \Sigma_n c_n o_n |\psi_n\rangle . \tag{4.6}$$

As explained in Section 2.5, forming the inner product of this new state $\widehat{O}|\psi\rangle$ with the original state $|\psi\rangle$ gives the expectation value of the observable corresponding to the operator. So in the case of observable O with associated operator $\widehat{O}$, eigenvalues o_n and expectation value $\langle O\rangle$,

$$\begin{aligned}\langle\psi|\widehat{O}|\psi\rangle &= \Sigma_m(c_m^* \langle\psi_m|)\Sigma_n(c_n o_n |\psi_n\rangle) \\ &= \Sigma_m\Sigma_n(c_m^* o_n c_n)\langle\psi_m|\psi_n\rangle = \Sigma_n o_n(|c_n|)^2 = \langle O\rangle ,\end{aligned}$$

since $\langle\psi_m|\psi_n\rangle = \delta_{m,n}$ for orthonormal wavefunctions. Note the role of the operator: it has performed the function of producing a new state in which the weighting coefficient of each eigenfunction has been multiplied by the eigenvalue of that eigenfunction. This is a crucial step in determining the expectation value of an observable.

The bottom line is this: applying an operator to the quantum state of a system changes that state by multiplying each constituent eigenfunction by its eigenvalue (Eq. 4.6), and making a measurement changes the quantum state by causing the wavefunction to collapse to one of those eigenfunctions. So operators and measurements both change the quantum state of a system, but not in the same way.

You can see examples of quantum operators in action later in this chapter and in Chapter 5, but before getting to that, you may find it helpful to consider the general characteristics of quantum wavefunctions, which is the subject of the next section.

4.3 Characteristics of Quantum Wavefunctions

In order to determine the details of the quantum wavefunctions that are solutions to the Schrödinger equation, you need to know the value of the potential energy V over the region of interest. You'll find solutions for several specific potentials in Chapter 5, but the general behavior of quantum wavefunctions can be discerned by considering the nature of the Schrödinger equation and the Copenhagen interpretation of its solutions.

For a function to qualify as a quantum wavefunction, it must be a solution to the Schrödinger equation, and it must also meet the requirements of the Born rule relating the squared magnitude of the function to the probability or probability density. Many authors of quantum texts describe such functions as "well-behaved," by which they usually mean that the function must be single-valued, smooth, and square-integrable. Here's a short explanation of what each of these terms means in this context and why these characteristics are required for quantum wavefunctions:

Single-valued This means that at any value of the argument of the function (such as x in the case of the wavefunction $\psi(x)$ in the one-dimensional position basis), the wavefunction can have only one value. That must be true for quantum-mechanical wavefunctions because the Born rule tells you that the square of the wavefunction gives the probability (or probability density in the case of continuous wavefunctions), which can have only one value at any location.

Smooth This means that the wavefunction and its first spatial derivative must be continuous, that is, with no gaps or discontinuities. That's because the Schrödinger equation is a second-order differential equation in the spatial

coordinate, and the second-order spatial derivative wouldn't exist if $\psi(x)$ or $\frac{\partial\psi(x)}{\partial x}$ were not continuous. An exception to this occurs in the case of infinite potential, which you can read about in the discussion of the infinite potential well in Chapter 5.

Square-integrable Quantum wavefunctions must be normalizable, which means that the integral of the wavefunction's squared magnitude cannot be infinitely large. For most functions, that means that the function itself must be finite everywhere, but you should be aware that the **Dirac delta function** is an exception. Although the Dirac delta function is defined to have infinite height, its infinitely narrow width keeps the area under its curve finite[3]. Note also that some functions such as the plane-wave function $Ae^{i(kx-\omega t)}$ have infinite spatial extent and are not individually square-integrable, but it is possible to construct combinations of these functions that have limited spatial extent and meet the requirement of square-integrability.

In addition to meeting these requirements, quantum wavefunctions must also match the boundary conditions for a particular problem. As mentioned earlier, it's necessary to know the specific potential $V(x)$ in the region of interest in order to fully determine the relevant quantum wavefunction. However, some important aspects of the wavefunction's behavior may be discerned by considering the relationship of wavefunction curvature to the value of the total energy E and the potential energy V in the time-independent Schrödinger equation (TISE).

To understand that behavior, it may help to do a bit of rearranging of the TISE (Eq. 3.40):

$$\frac{d^2[\psi(x)]}{dx^2} = -\frac{2m}{\hbar^2}(E-V)\psi(x). \tag{4.7}$$

The term on the left side of this equation is just the spatial curvature, which is the change in the slope of a graph of the $\psi(x)$ function vs. location. According to this equation, that curvature is proportional to the wavefunction $\psi(x)$ itself, and one of the factors of proportionality is the quantity $E - V$, the difference between the total energy and the potential energy at the location under consideration.

[3] Not technically a function due to its infinitely large value when its argument is zero, the Dirac delta function is actually a "generalized function" or "**distribution**," which is the mathematical equivalent of a black box that produces a known output for a given input. In physics, the usefulness of the Dirac delta function is usually realized when it appears inside an integral, as you can see in the discussion of Fourier analysis in Section 4.4.

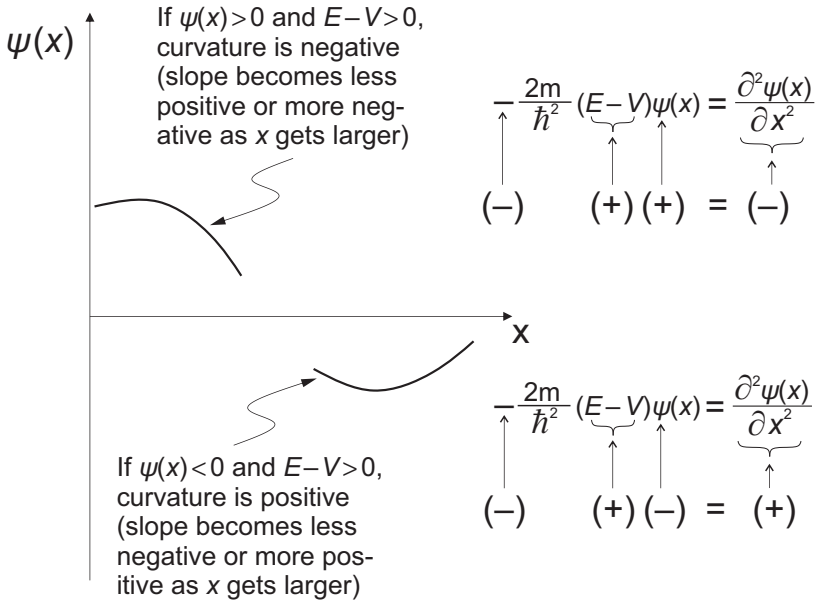

ψ(x)
If ψ(x)>0 and E−V>0, curvature is negative (slope becomes less positive or more negative as x gets larger)
x
If ψ(x)<0 and E−V>0, curvature is positive (slope becomes less negative or more positive as x gets larger)
−2m/ħ² (E−V)ψ(x) = ∂²ψ(x)/∂x²
(−) (+) (+) = (−)
−2m/ħ² (E−V)ψ(x) = ∂²ψ(x)/∂x²
(−) (+) (−) = (+)

Figure 4.1 Wavefunction curvature for $E - V > 0$ case.

Now imagine a situation in which the total energy (E) exceeds the potential energy (V) everywhere in the region of interest (this doesn't necessarily imply that the potential energy is constant, just that the total energy is greater than the potential energy at every location in this region). Hence $E - V$ is positive, and the curvature has the opposite sign of the wavefunction (due to the minus sign on the right side of Eq. 4.7).

Why do the signs of the curvature and the wavefunction matter? Look at the behavior of the wavefunction as you move toward positive x in a region in which the wavefunction $\psi(x)$ is positive (above the x-axis in Fig. 4.1). Since $E - V$ is positive, the curvature of $\psi(x)$ must be negative in this region (since the curvature and the wavefunction have the opposite sign if $E - V$ is positive). This means that the slope of the graph of $\psi(x)$ becomes increasingly negative as you move to larger values of x, so the waveform must curve *toward* the x-axis, eventually crossing that axis. When that happens, the wavefunction $\psi(x)$ becomes negative, and the curvature becomes positive. That means that the waveform again curves toward the x-axis, until it eventually crosses back into the positive-ψ region, where the curvature once again becomes negative.

So no matter how the potential energy function $V(x)$ behaves, as long as the total energy exceeds the potential energy, the wavefunction $\psi(x)$ will oscillate as a function of position. As you'll see later in this chapter and in Chapter 5, the wavelength and amplitude of those oscillations are determined by the difference between E and V.

Now consider a region in which the total energy is less than the potential energy. That means that $E - V$ is negative, so the curvature has the same sign as the wavefunction.

If you're new to quantum mechanics, the idea of the total energy being less than the potential energy may seem to be physically impossible. After all, if the total energy is the sum of potential plus kinetic energy, wouldn't the kinetic energy have to be negative to make the total energy less than the potential energy? And how can the kinetic energy, which is $\frac{1}{2}mv^2 = \frac{p^2}{2m}$, be negative?

In classical physics, that line of reasoning is correct, which is why regions in which an object's potential energy exceeds its total energy are called "**classically forbidden**" or "classically disallowed." But in quantum mechanics, solving the Schrödinger equation in a region in which the potential energy exceeds the total energy leads to perfectly acceptable wavefunctions; as you'll see later in the section, those wavefunctions decay exponentially with distance within those regions. And what happens if you measure the kinetic energy in one of these regions? The wavefunction will collapse to one of the eigenstates of the kinetic-energy operator, and the result of your measurement will be the eigenvalue of that eigenstate. Those eigenvalues are all positive, so you will definitely not measure a negative value for kinetic energy.

How is that result consistent with the potential energy exceeding the total energy in this region? The answer is hidden in the phrase "in this region." Since position and momentum are incompatible observables, and kinetic energy depends on the square of momentum, the uncertainty principle says that you cannot simultaneously measure both position and kinetic energy with arbitrarily great precision. Specifically, the more precisely you measure kinetic energy, the larger the uncertainty in position. So when you measure kinetic energy and get a positive value, the possible positions of the quantum particle or system always include a region in which the total energy exceeds the potential energy.

With that understanding, take a look at the behavior of the wavefunction $\psi(x)$ as you move toward positive x in a region in which $\psi(x)$ is initially positive (above the x-axis) in Fig. 4.2. If $E - V$ is negative in this region, then the curvature must be positive, since the curvature and the wavefunction must have the same sign if $E - V$ is negative and $\psi(x)$ is positive. This means that the slope of the graph of $\psi(x)$ becomes increasingly positive as you move

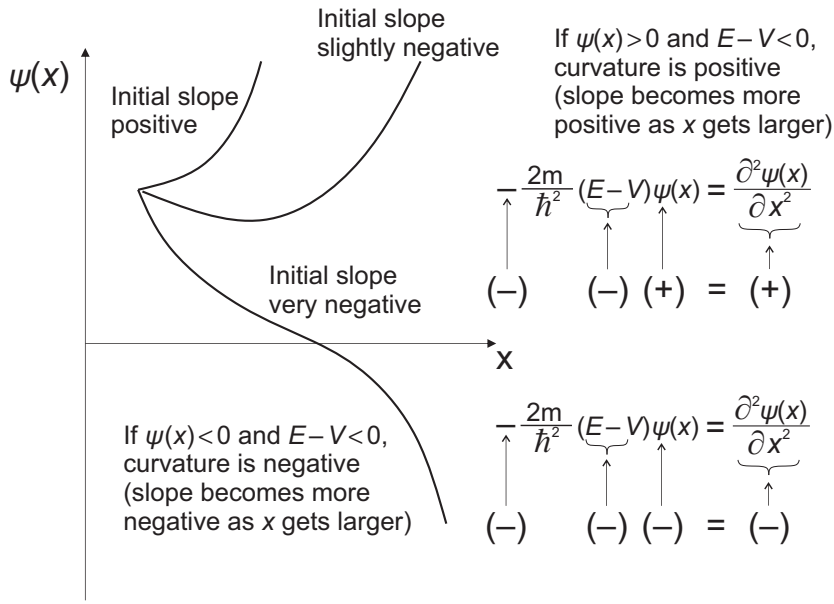

ψ(x)
Initial slope slightly negative
Initial slope positive
If ψ(x)>0 and E−V<0, curvature is positive (slope becomes more positive as x gets larger)
Initial slope very negative
x
If ψ(x)<0 and E−V<0, curvature is negative (slope becomes more negative as x gets larger)

Figure 4.2 Wavefunction curvature for $E - V < 0$ case.

to larger values of x, which means that the waveform must curve *away* from the x-axis. And if the slope of $\psi(x)$ is positive (or zero) at the position shown, the wavefunction will eventually become infinitely large.

Now think about what happens if the slope of $\psi(x)$ is negative at the position shown. That depends on exactly *how negative* that slope is. As shown in the figure, even if the slope is initially slightly negative, the positive curvature will cause the slope to become positive, and the graph of $\psi(x)$ will turn upward before crossing the x-axis, which means the value of $\psi(x)$ will eventually become infinitely large.

But if the slope at the position shown is sufficiently negative, $\psi(x)$ will cross the x-axis and become negative. And when $\psi(x)$ becomes negative, the curvature will also become negative, since $E - V$ is negative. With negative curvature below the axis, $\psi(x)$ will curve away from the x-axis, eventually becoming infinitely large in the negative direction.

So for each of the initial slopes shown in Fig. 4.2, the value of the wavefunction $\psi(x)$ will eventually reach either $+\infty$ or $-\infty$. And since wavefunctions with infinitely large amplitude are not physically realizable, the slope of the wavefunction cannot have any of these values at the position

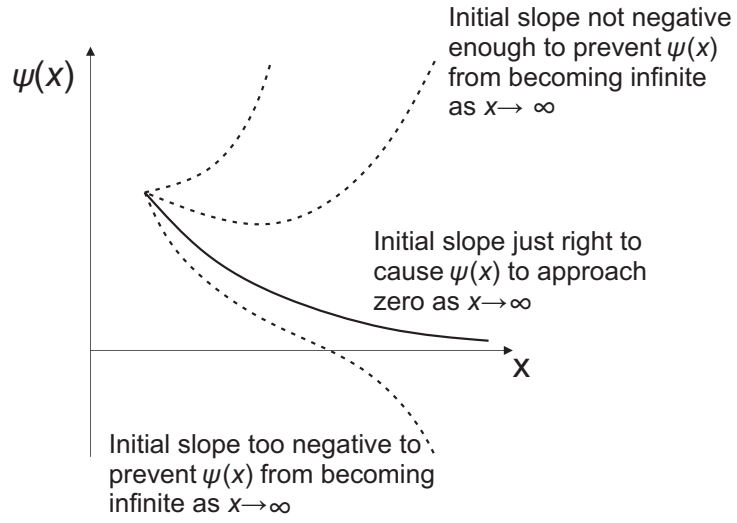

ψ(x)
Initial slope not negative enough to prevent ψ(x) from becoming infinite as x→ ∞
Initial slope just right to cause ψ(x) to approach zero as x→∞
x
Initial slope too negative to prevent ψ(x) from becoming infinite as x→∞

Figure 4.3 Effect of initial slope on $\psi(x)$ for $E - V < 0$ case.

shown. Instead, the curvature of the wavefunction must be such that the amplitude of $\psi(x)$ remains finite at all locations so that the wavefunction is normalizable. This means that the integral of the square magnitude of $\psi(x)$ must converge to a finite value, which means that the value of $\psi(x)$ must tend toward zero as x approaches $\pm\infty$. For that to happen, the slope $\frac{\partial\psi}{\partial x}$ must have just the right value to cause $\psi(x)$ to approach the x-axis asymptotically, never turning away from the axis, but also never crossing below the axis. In that case, $\psi(x)$ will approach zero as x approaches ∞, as shown in Fig. 4.3.

What conclusion can you reach about the behavior of the wavefunction $\psi(x)$ in the regions in which $E - V$ is negative? Just this: oscillations are not possible in such regions, because the slope of the wavefunction at any position must have the value that will cause the wavefunction to decay toward zero as x approaches $\pm\infty$.

So just by considering the Schrödinger equation's relationship of wavefunction curvature to the value of $E - V$, you can determine that $\psi(x)$ oscillates in regions in which the total energy E exceeds the potential energy V and decays in regions in which E is less than V. More details of that behavior can be found by solving the Schrödinger equation for specific potentials, as you can see in Chapter 5.

To get a better sense of how the wavefunction behaves, consider the cases in which the potential $V(x)$ is constant over the region of interest and the total energy E is either greater than or less than the potential energy.

First taking the case in which $E - V$ is positive, the TISE (Eq. 4.7) can be written as

$$\frac{d^2[\psi(x)]}{dx^2} = -\frac{2m}{\hbar^2}(E - V)\psi(x) = -k^2\psi(x), \tag{4.8}$$

in which the constant k is given by

$$k = \sqrt{\frac{2m}{\hbar^2}(E - V)}. \tag{4.9}$$

The general solution to this equation is

$$\psi(x) = Ae^{ikx} + Be^{-ikx}, \tag{4.10}$$

in which A and B are constants to be determined by the boundary conditions.[4]

Even without knowing those boundary conditions, you can see that quantum wavefunctions oscillate sinusoidally in regions in which E is greater than V (classically allowed regions), since $e^{\pm ikx} = \cos kx \pm i \sin kx$ by Euler's relation. This fits with the curvature analysis presented earlier.

And here's another conclusion you can draw from the form of the solution in Eq. 4.10: k represents the wavenumber in this region, which determines the wavelength of the quantum wavefunction through the relation $k = 2\pi/\lambda$. The wavenumber determines how "fast" the wavefunction oscillates with distance (cycles per meter rather than cycles per second), and Eq. 4.9 tells you that large $E - V$ means large k, and large wavenumber means short wavelength. Thus the larger the difference between the total energy E and the potential energy V of a quantum particle, the greater the curvature and the faster the particle's wavefunction will oscillate with x (higher number of cycles per meter).

Enforcing the boundary conditions of continuous $\psi(x)$ and continuous slope ($\frac{\partial\psi(x)}{\partial x}$) at the boundary between two classically allowed regions with different potentials can help you understand the relative amplitude of the wavefunction in the two regions. To see how that works, use Eq. 4.10 to write the wavefunction and its first spatial derivative on both sides of the boundary. Since taking that derivative brings out a factor of k, this leads to the conclusion that the ratio of the amplitudes on opposite sides of the boundary is inversely

[4] Note that this is equivalent to $A_1 \cos(kx) + B_1 \sin(kx)$ and to $A_2 \sin(kx + \phi)$; you can see why that's true as well as the relationship between the coefficients of these equivalent expressions in the chapter-end problems and online solutions.

proportional to the wavenumber ratio.[5] Thus the wavefunction on the side of the boundary with larger energy difference $E - V$ (which means larger k) must have smaller amplitude than the wavefunction on the opposite side of the boundary between classically allowed regions.

Now consider the case in which the potential energy exceeds the total energy, so $E - V$ is negative. In that case, the TISE (Eq. 4.7) can be written as

$$\frac{d^2[\psi(x)]}{dx^2} = -\frac{2m}{\hbar^2}(E - V)\psi(x) = +\kappa^2\psi(x), \tag{4.11}$$

in which the constant κ is given by

$$\kappa = \sqrt{\frac{2m}{\hbar^2}(V - E)}. \tag{4.12}$$

The general solution to this equation is

$$\psi(x) = Ce^{\kappa x} + De^{-\kappa x}, \tag{4.13}$$

in which C and D are constants to be determined by the boundary conditions.

If the region of interest is a classically forbidden region extending toward $+\infty$ (so that x can take on large positive values within the region), the first term of Eq. 4.13 will become infinitely large unless the coefficient C is set to zero. In this region, $\psi(x) = De^{-\kappa x}$, which decreases exponentially with increasing positive x.

Likewise, if the region of interest is a classically forbidden region extending toward $-\infty$, the second term of Eq. 4.13 will become infinitely large as x takes on large negative values, so in that case the coefficient D must be zero. That makes $\psi(x) = Ce^{\kappa x}$ in this region, and the wavefunction amplitude decreases exponentially with increasing negative x.

So once again, even without knowing the precise boundary conditions, you can conclude that quantum wavefunctions decay exponentially in regions in which E is greater than V (that is, in classically forbidden regions), again in accordance with the curvature analysis presented earlier.

Additional information can be gleaned from Eq. 4.13: the constant κ is a "decay constant" that determines the rate at which the wavefunction tends toward zero. And since Eq. 4.12 states that κ is directly proportional to the square root of $V - E$, you know that the greater the amount by which the potential energy V exceeds the total energy E, the larger the decay constant κ, and the faster the wavefunction decays with increasing x.

[5] If you need help getting that result, check out the chapter-end problems and online solutions.

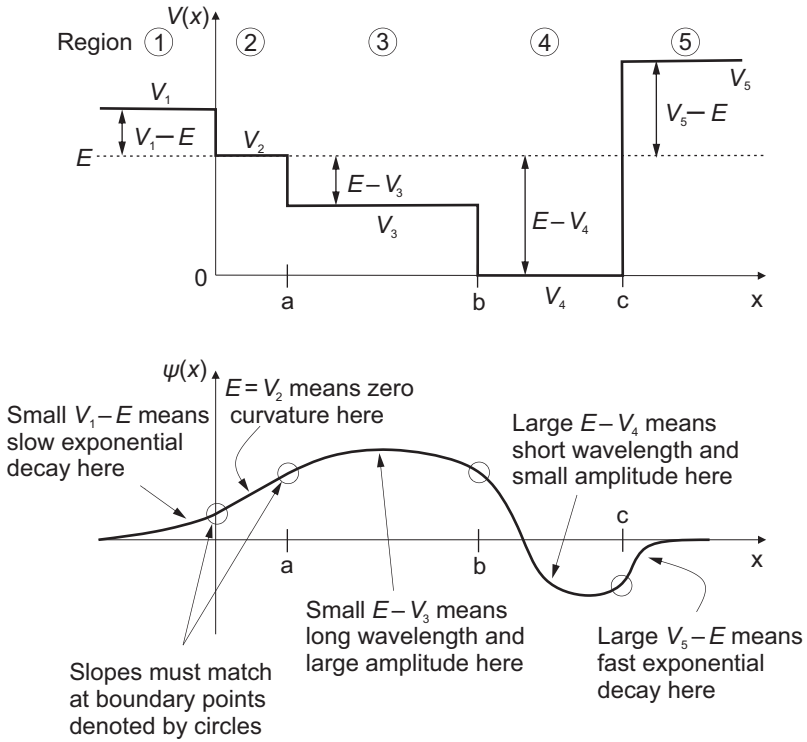


Figure 4.4 Stepped potential and wavefunction.

You can see all of these characteristics at work in Fig. 4.4, which shows five regions over which the potential has different values (but $V(x)$ is constant within each region). Such "piecewise constant" potentials are helpful for understanding the behavior of quantum wavefunctions, and they can also be useful for simulating continuously varying potential.

The potential V_1 in region 1 (leftmost) and the potential V_5 in region 5 (rightmost) are both greater than the particle's energy E, albeit by different amounts. In regions 2, 3, and 4, the particle's energy is greater than the potential, again by a different amount in each region.

In classically forbidden regions 1 and 5, the wavefunction decays exponentially, and since $V - E$ is greater in region 5 than in region 1, the decay over distance is faster in that region.

In classically allowed region 2, the total energy and the potential energy are equal, so the curvature is zero in that region. Note also that the slope of the

wavefunction is continuous across the boundary between classically forbidden region 1 and allowed region 2.

In the classically allowed regions 3 and 4, the wavefunction oscillates, and since the difference between the total and potential energy is smaller in region 3, the wavenumber k is smaller in that region, which means that the wavelength is longer and the amplitude is larger. The larger value of $E - V$ in region 4 makes for shorter wavelength and smaller amplitude in that region.

At each boundary between two regions (marked by circles in Fig. 4.4), whether classically allowed or forbidden, both the wavefunction $\psi(x)$ and the slope $\frac{\partial \psi}{\partial x}$ must be continuous (that is, the same on both sides of the boundary).

Another aspect of the potentials and wavefunction shown in Fig. 4.4 is worth considering: for a particle with the total energy E shown in the figure, the probability of finding the particle decreases to zero as x approaches $\pm\infty$. That means that the particle is in a bound state – that is, localized to certain region of space. Unlike such bound particles, free particles are able to "escape to infinity" in the sense that their wavefunctions are oscillatory over all space. As you'll see in Chapter 5, particles in bound states have a discrete spectrum of allowed energies, while free particles have a continuous energy spectrum.

4.4 Fourier Theory and Quantum Wave Packets

If you've worked through the previous chapters, you've briefly encountered the two major aspects of Fourier theory: analysis and synthesis. In Section 1.6, you learned how to use the inner product to find the components of a wavefunction, and Fourier analysis is one type of that "spectral decomposition." In Section 3.1, you saw how to produce a composite wavefunction by the weighted addition of plane-wave functions, and superposition of sinusoidal functions is the basis of Fourier synthesis.

The goal of this section is to help you understand why the Fourier transform plays a key role in both analysis and synthesis, and exactly how that transform works. You'll also see how Fourier theory can be used to understand quantum **wave packets** and how it relates to the uncertainty principle.

To understand exactly what the Fourier transform does to a function, consider the mathematical statement of the Fourier transform of a function of position $\psi(x)$:

$$\phi(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \psi(x) e^{-ikx} dx, \tag{4.14}$$

in which $\phi(k)$ is a function of wavenumber (k) called the wavenumber spectrum.

If you already know the wavenumber spectrum $\phi(k)$ and you want to determine the corresponding position function $\psi(x)$, the tool you need is the inverse Fourier transform

$$\psi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \phi(k) e^{ikx} dk. \tag{4.15}$$

Fourier theory (both analysis and synthesis) is rooted in one idea: any well-behaved[6] function can be expressed as a weighted combination of sinusoidal functions. In the case of a function of position such as $\psi(x)$, the constituent sinusoidal functions are of the form $\cos kx$ and $\sin kx$, with k representing the wavenumber of each component (recall that wavenumber is sometimes called "spatial frequency" and has dimensions of angle per unit length, with SI units of radians per meter).

To understand the meaning of the Fourier transform, imagine that you have a function of position $\psi(x)$ and you want to know "how much" of each constituent cosine and sine function is present in $\psi(x)$ for each wavenumber k. What Eq. 4.14 is telling you is this: to find those amounts, multiply $\psi(x)$ by e^{-ikx} (which is equivalent to $\cos kx - i \sin kx$ by Euler's relation) and integrate the resulting product over all space. The result of that process is the complex function $\phi(k)$. If $\psi(x)$ is real, then the real part of $\phi(k)$ tells you the amount of $\cos kx$ in $\psi(x)$ and the imaginary part of $\phi(k)$ tells you the amount of $\sin kx$ present in $\psi(x)$ for each value of k.

Why does this process of multiplying and integrating tell you the amount of each sinusoid in the function $\psi(x)$? There are several ways to picture this, but some students find that the simplest visualization comes from using the Euler relation to write the Fourier transform as

$$\begin{aligned} \phi(k) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \psi(x) e^{-ikx} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \psi(x) \cos(kx) dx - i \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \psi(x) \sin(kx) dx. \end{aligned} \tag{4.16}$$

Now imagine the case in which the function $\psi(x)$ is a cosine function with single wavenumber k_1, so $\psi(x) = \cos(k_1 x)$. Inserting this into Eq. 4.16 gives

$$\phi(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \cos(k_1 x) \cos(kx) dx - i \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \cos(k_1 x) \sin(kx) dx.$$

[6] In this context, "well-behaved" means that the function satisfies the **Dirichlet conditions** of finite number of extrema and finite number of non-infinite discontinuities.

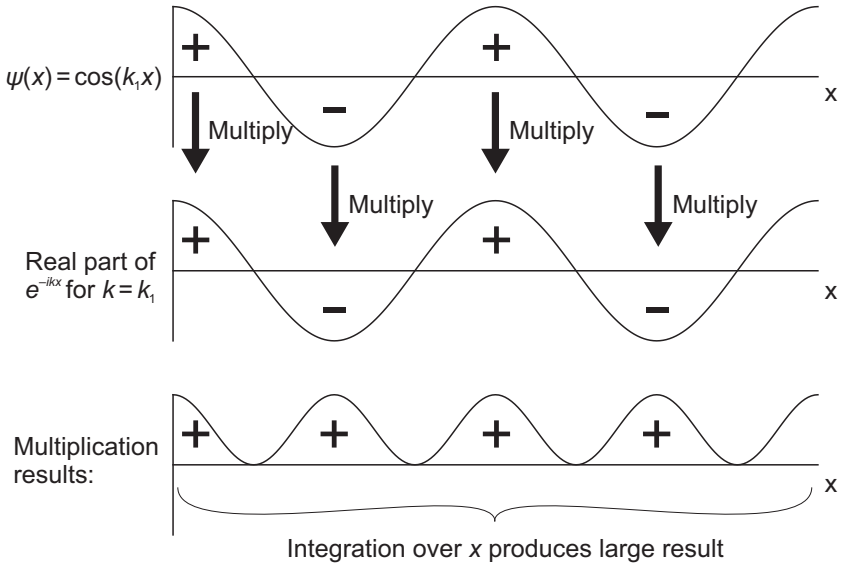


Figure 4.5 Multiplying and integrating $\cos k_1 x$ and the real portion of e^{-ikx} when $k = k_1$.

The next step in many explanations is to invoke the "**orthogonality relations**," which say that the first integral is nonzero only when $k = k_1$ (since cosine waves with different spatial frequencies are orthogonal to one another when integrated over all space), while the second integral is zero for all values of k (since the sine and cosine functions are also orthogonal when integrated over all space). But if you're not clear on why that's true, take a look at Fig. 4.5, which provides more detail about the orthogonality of sinusoidal functions described in Section 1.5.

The top graph in this figure shows the single-wavenumber wavefunction $\psi(x) = \cos(k_1 x)$ and the center graph shows the real portion of the function e^{-ikx} (which is $\cos kx$) for the case in which $k = k_1$. The vertical arrows indicate the point-by-point multiplication of these two functions, and the bottom portion of the figure shows the result of that multiplication. As you can see, since all of the positive and negative portions of $\psi(x)$ align with the portions of the real part of e^{-ikx} with the same sign, the results of the multiplication process are all positive (albeit with varying amplitude due to the oscillations of both functions, as shown in the bottom graph). Integrating the multiplication product over x is equivalent to finding the area under this

curve, and that area will have a large value when the product always has the same sign. In fact, the area under the curve is infinite if the integration extends from $-\infty$ to $+\infty$ and if the products are all positive, which means that k is precisely equal to k_1. But even the slightest difference between k and k_1 will cause the two functions to go from in-phase to out-of-phase and back to in-phase at a rate determined by the difference between k and k_1, which will cause the product of $\cos kx$ and $\cos k_1 x$ to oscillate between positive and negative values. That means that the result of integration over all space approaches infinity for $k = k_1$ and zero for all other values of k, resulting in a function that's infinitely tall but infinitely narrow. That function is the Dirac delta function, about which you can read more later in this section.

So $\phi(k)$, the Fourier transform of the constant-amplitude (and thus infinitely wide) wavefunction $\psi(x) = \cos k_1 x$ has an infinitely large real value at wavenumber k_1. What about the imaginary portion of $\phi(k)$? Since $\psi(x)$ is a pure (real) cosine wave, you can probably guess that no sine function is included in $\psi(x)$, even at the wavenumber k_1. That's exactly what the Fourier transform produces, as shown in Fig. 4.6.

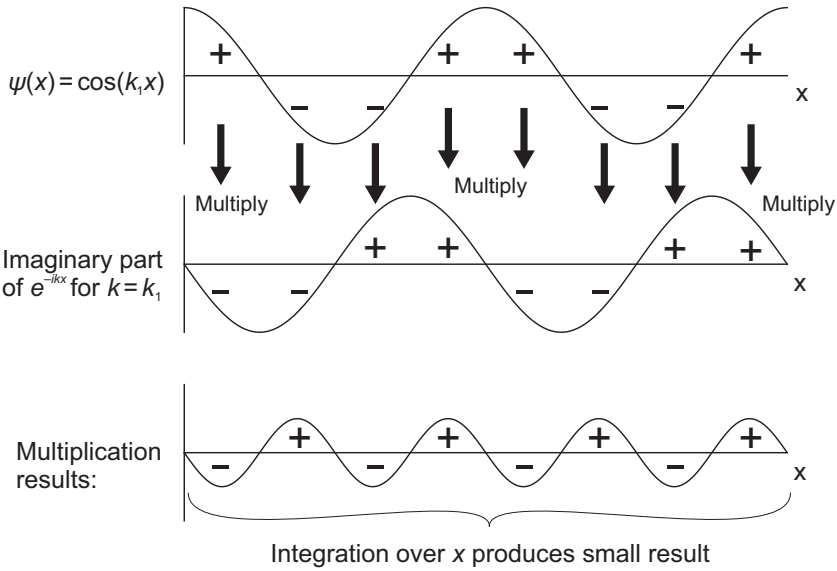


Figure 4.6 Multiplying and integrating $\cos k_1 x$ and the imaginary portion of e^{-ikx} when $k = k_1$.

Notice that even though the oscillations of $\psi(x)$ and the imaginary portion of e^{-ikx} (which is $-\sin kx$) have the same spatial frequency in this case, the phase offset between these two functions makes their product equally positive and negative. Hence integrating over x produces a small result (zero if the integration is done over an integer number of cycles, as explained later in this section). So $\phi(k)$ will have small or zero imaginary portion, even at wavenumber $k = k_1$.

Since $\psi(x)$ is a pure cosine wave in this example, will the Fourier transform produce precisely zero for the imaginary portion of $\phi(k)$? It will if you integrate over an integer number of cycles of $\psi(x)$, because in that case the result of the multiplication will have exactly as much positive as negative contribution to $\phi(k)$ (that is, the area under the curve will be exactly zero). But if you integrate over, say, 1.25 cycles of $\psi(x)$, there will be some residual negative area under the curve, so the result of the integration will not be exactly zero. Note, however, that you can make the ratio of the imaginary part to the real part arbitrarily small by integrating over the entire x-axis, since that will cause the real portion of $\phi(k)$ (with its all-positive multiplication results) to greatly exceed any unbalanced positive or negative area in the imaginary portion. That's why the limits of integration on Fourier orthogonality relations are $-\infty$ to ∞ in the general case, or $-T/2$ to $T/2$ for periodic functions (where T is the period of the function being analyzed).

So the multiply-and-integrate process of the Fourier transform produces the expected result when the wavenumber k of the e^{-ikx} factor matches the wavenumber of one of the components of the function being transformed (k_1 in this case). But what happens at other values of the wavenumber k? Why does this process lead to small values of $\phi(k)$ for wavenumbers that are not present in $\psi(x)$?

To understand the answer to that question, consider the multiplication results shown in Fig. 4.7. In this case, the wavenumber k in the multiplying factor e^{-ikx} is taken to be half the value of the single wavenumber (k_1) in $\psi(x)$. As you can see in the figure, in this case each spatial oscillation of the real portion of the $e^{-i(\frac{1}{2})k_1x}$ factor occurs over twice the distance of each oscillation of $\psi(x)$. That means that the product of these two functions alternates between positive and negative, making the area under the resulting curve tend toward zero. The changing amplitudes of $\psi(x)$ and $e^{-i(\frac{1}{2})k_1x}$ cause the amplitude of their product to vary over x, but the symmetry of the waveforms ensures the equality of the positive and negative areas over any integer number of cycles.

A similar analysis shows that the Fourier transform also produces a small result when the wavenumber in the multiplying factor e^{-ikx} is taken to be

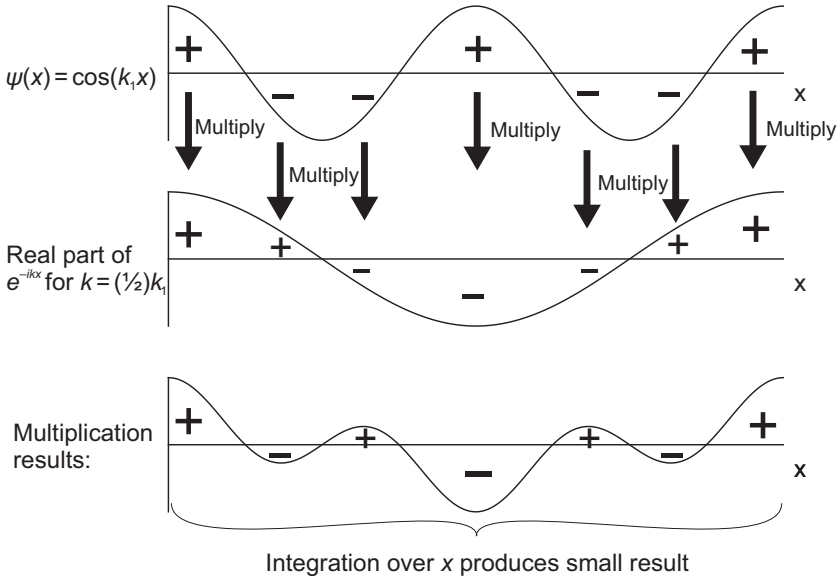


Figure 4.7 Multiplying and integrating $\cos k_1 x$ and the imaginary portion of e^{-ikx} when $k = \frac{1}{2}k_1$.

larger than the value of the single wavenumber (k_1) in $\psi(x)$. Fig. 4.8 shows what happens for the case in which $k = 2k_1$, so each spatial oscillation of the $e^{-i(2)k_1 x}$ factor occurs over half the distance of each oscillation of $\psi(x)$. As in the previous case, the product of these two functions alternates between positive and negative, and once again the area under the resulting curve tends toward zero (and is precisely zero over any integer number of cycles of $\psi(x)$).

You should make sure you understand that in this example, the imaginary portion of $\phi(k)$ is zero because $\psi(x)$ is a pure cosine wave, not because $\psi(x)$ is real. If $\psi(x)$ had been a pure (real) sine wave, the result $\phi(k)$ of the Fourier transform process would have been purely imaginary, because in that case only one sine and no cosine components are needed to make up $\psi(x)$. In general, the result of the Fourier transform is complex, whether the function you're transforming is purely real, purely imaginary, or complex.

Multiplying the function being transformed by the real and imaginary parts of e^{-ikx} is one way to understand the process of Fourier transformation, but there's another way that has some benefits due to the complex nature of $\psi(x)$, $\phi(k)$, and the transformation process. That alternative approach is to represent the components of $\psi(x)$ and the multiplying factor e^{-ikx} as phasors.

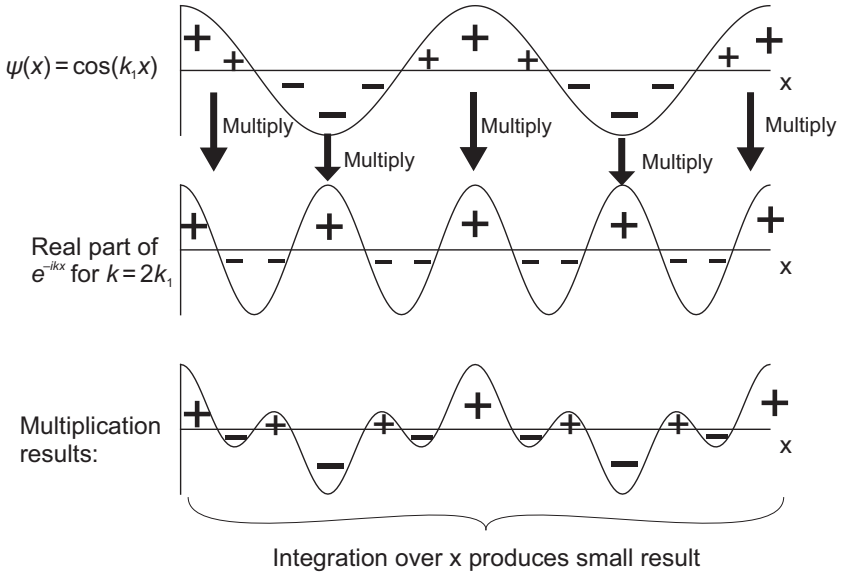


Figure 4.8 Multiplying and integrating $\cos k_1 x$ and the imaginary portion of e^{-ikx} when $k = 2k_1$.

If it's been a while since you've seen phasors, and even if you never really understood them, don't worry. Phasors are a very convenient way to represent the sinusoidal functions at the heart of the Fourier transform, and the next few paragraphs will provide a quick review of phasor basics if you need it.

Phasors dwell in the complex plane described in Section 1.4. Recall that the complex plane is a two-dimensional space defined by a "real" axis (usually horizontal) and a perpendicular "imaginary" axis (usually vertical). A phasor is a type of vector in that plane, typically drawn with its base at the origin and its tip on the unit circle, which is the locus of points at a distance of one unit from the origin.

The reason that phasors are helpful in representing sinusoidal functions is this: by making the angle of the phasor from the positive real axis equal to kx (proceeding counterclockwise as kx increases), the functions $\cos(kx)$ and $\sin(kx)$ are traced out by projecting the phasor onto the real and imaginary axes. You can see this in Fig. 4.9, in which the rotating phasor is shown at eight randomly selected values of the angle kx. The phasor rotates continuously as x increases, with the rate of rotation determined by the wavenumber k. Since

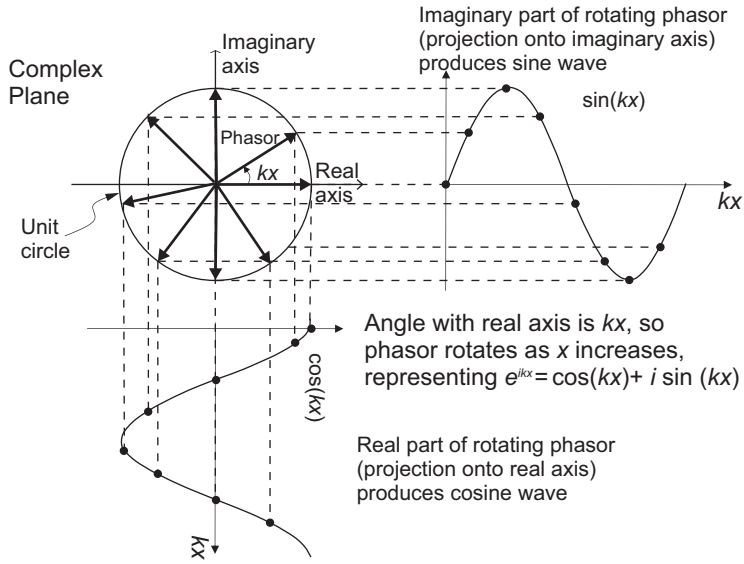


Figure 4.9 Rotating phase relation to sine and cosine functions.

$k = \frac{2\pi}{\lambda}$, an increase in the value of x of one wavelength causes kx to increase by 2π radians, which means the phasor will make one complete revolution.

To use phasors to understand the Fourier transform, imagine that the function $\psi(x)$ under analysis has a single, complex wavenumber component represented by e^{ik_1x}. The phasor representing $\psi(x)$ is shown in Fig. 4.10a at 10 evenly spaced values of k_1x, which means that the value of x is increasing by $\frac{\lambda}{10}$ (angle of $\frac{\pi}{5}$ radians or 36°) between each position of the phasor.

Now look at Fig. 4.10b, which shows the phasor representing the Fourier-transform multiplying factor e^{-ikx} for the case in which $k = k_1$. The phasor representing this function rotates at the same rate as the phasor representing $\psi(x)$, but the negative sign in the exponent of this function means that its phasor rotates clockwise. To appreciate what this means for the result of the Fourier-transform process, it's important to understand the effect of multiplying two phasors (such as $\psi(x)$ times e^{-ikx}).

Multiplying two phasors produces another phasor, and the amplitude of that new phasor is equal to the product of the amplitudes of the two phasors (both equal to one if their tips lie on the unit circle). More importantly for this application, the direction of the new phasor is equal to the sum of the angles of the two phasors being multiplied. And since the two phasors in this case

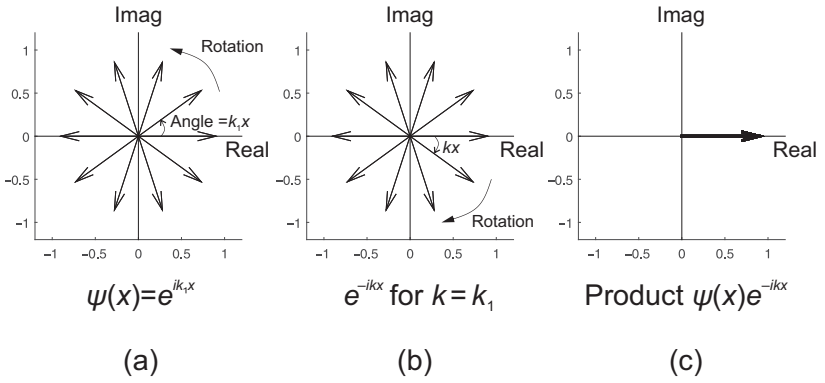


Figure 4.10 Phasor representation of (a) the function e^{ik_1x}, the multiplying factor e^{-ikx} for $k = k_1$, and (c) the product.

are rotating in opposite directions at the same rate, the sum of their angles is constant. To see why that's true, begin by defining the direction of the real axis to represent 0°. The positions shown for the $\psi(x)$ phasor are then 36°, 72°, 108°, and so on, while the positions shown for the clockwise-rotating phasor representing e^{-ikx} are −36°, −72°, −108°. That makes the sum of the two phasors' angles equal to zero for all values of x.[7] Hence the multiplications performed at the ten phasor angles shown in Fig. 4.10a all result in a unit-length phasor pointing along the real axis, as shown in Fig. 4.10c.

Why is it significant that the phasor resulting from the multiplication of the phasors representing $\psi(x)$ and e^{-ikx} has constant direction? Because the Fourier-transform process integrates the result of that multiplication over all values of x:

$$\phi(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \psi(x) e^{-ikx} dx, \tag{4.14}$$

which is equivalent to continuously adding the resulting phasors. Since those phasors all point in the same direction when $k = k_1$ (that is, when $\psi(x)$ contains a wavenumber component that matches the wavenumber in the multiplying function e^{-ikx}), that addition yields a large number (phasors follow the rules of vector addition, so the largest possible sum occurs when they all point in the same direction). That large number becomes infinite when the

[7] If you prefer to use positive angles, the clockwise-rotating phasor's angles count down from 360° to 324°, 288°, 252°, and so on, which sum with the angles of the $\psi(x)$ phasor to give a constant value of 360°, the same direction as 0°.

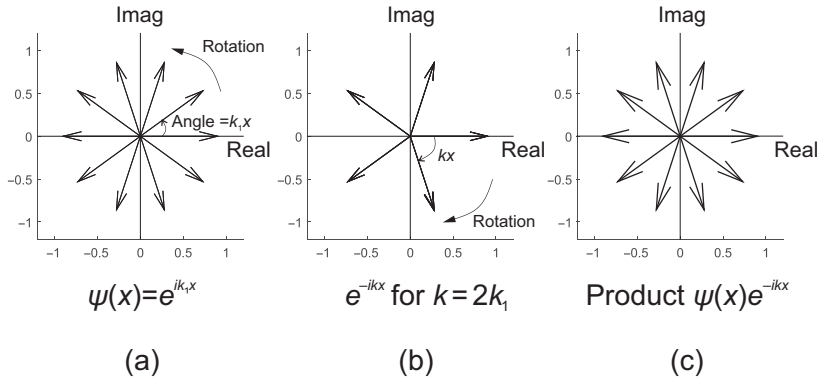


Figure 4.11 Phasor representation of (a) the function e^{ik_1x}, (b) the multiplying factor e^{-ikx} for $k = 2k_1$, and (c) their product.

value of k precisely equals the value of k_1 and the integration extends from $-\infty$ to $+\infty$.

The situation is quite different when a wavenumber contained within $\psi(x)$ (such as k_1) does not match the wavenumber of the multiplying function e^{-ikx}). An example of that is shown in Fig. 4.11, for which the k in e^{-ikx} is twice the value of the wavenumber k_1 present in $\psi(x)$. As you can see in Fig. 4.11b, the larger wavenumber causes the phasor representing this function to rotate at a higher angular rate. That rate is twice as large in this case, so this phasor advances 72° as the $\psi(x)$ phasor advances 36°, and it completes two cycles as the $\psi(x)$ phasor shown in Fig. 4.11a completes one.

The important consequence of these different angular rates is that the angle of the phasor produced by multiplying $\psi(x)$ by e^{-ikx} is not constant. With both phasors starting at 0° when $x = 0$, after one increment the $\psi(x)$ phasor's angle is 36°, while the e^{-ikx} phasor's angle is −72°, so their product phasor's angle is $36° + (-72°) = -36°$. As x increases one more of these increments, the $\psi(x)$ phasor's angle becomes 72°, while the e^{-ikx} phasor's angle becomes −144°, making their product phasor's angle $72° + (-144°) = -72°$. When the increase in x has caused the $\psi(x)$ phasor to complete one revolution (and the e^{-ikx} phasor to complete two revolutions), their product phasor will also have completed one clockwise cycle, as shown in Fig. 4.11c.

The changing angles of those phasors means that their sum will tend toward zero, as you can determine by lining them up in the head-to-tail configuration of vector addition (since they will form a loop and end up back at the starting point). And the result of integrating the product of $\psi(x)$ and e^{-ikx} will be

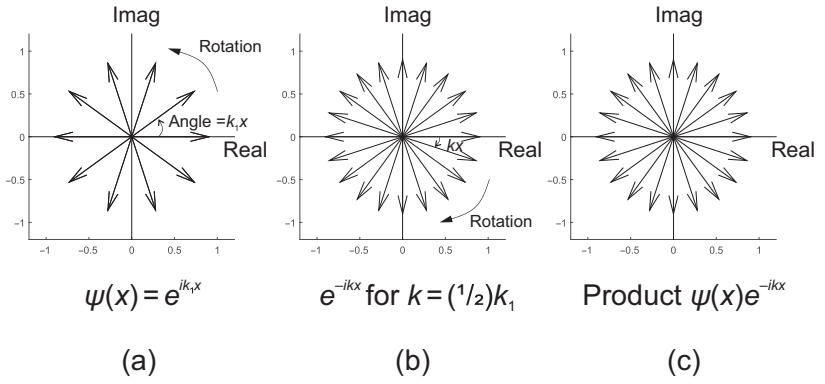


Figure 4.12 Phasor representation of (a) the function e^{ik_1x}, (b) the multiplying factor e^{-ikx} for $k = \frac{1}{2}k_1$, and (c) their product.

exactly zero if the integration is performed over a sufficiently large range of x to cause the product phasor to complete an integer number of cycles.

As you may have guessed, a similar analysis applies when the wavenumber component in the function e^{-ikx} is smaller than the wavenumber k_1 in $\psi(x)$. The case for which $k = \frac{1}{2}k_1$ is shown in Fig. 4.12; you can see the phasor representing e^{-ikx} taking smaller steps (18° in this case) in Fig. 4.12b. And since k does not match k_1, the direction of the product phasor is not constant; in this case it rotates counter-clockwise. But as long as the product phasor completes an integer number of cycles in either direction, the integration of the products will give zero.

You can use this same type of phasor analysis for functions that cannot be represented by a single rotating phasor with constant amplitude. For example, consider the wavefunction $\psi(x) = \cos(k_1x)$ discussed earlier in the section. Since this is a purely real function, a single rotating phasor won't do. But recall the "inverse Euler" relation for cosine:

$$\cos(k_1x) = \frac{e^{ik_1x} + e^{-ik_1x}}{2}. \tag{4.17}$$

This means that the function $\psi(x) = \cos(k_1x)$ may be represented by two counter-rotating phasors with amplitude of $\frac{1}{2}$, as shown in Fig. 4.13a. As these two phasors rotate in opposite directions, their sum (not their product) lies entirely along the real axis (since their imaginary components have opposite signs, and cancel). Over each complete rotation of the two component phasors, the amplitude of the resultant phasor varies from +1, through 0 (when they

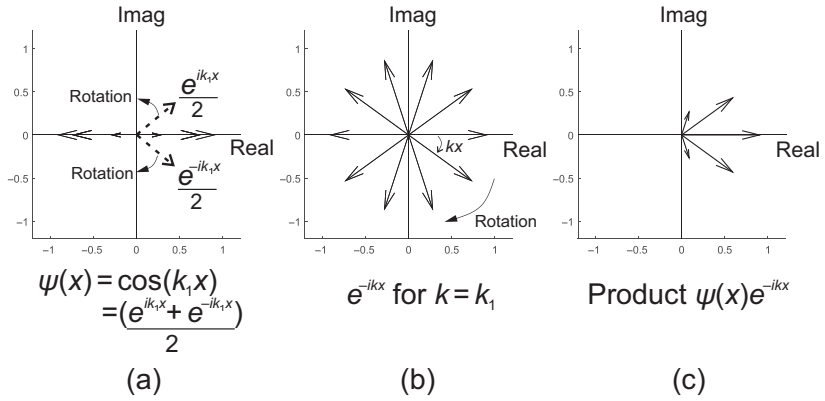


Figure 4.13 Phasor representation of (a) the function $\cos(k_1 x)$, (b) the multiplying factor e^{-ikx} for $k = k_1$, and (c) their product.

point in opposite directions along the imaginary axis), to -1, through 0 again, and back to -1, exactly as a cosine function should.

You can see how the phasor analysis of the Fourier transform works in this case by looking at Fig. 4.13b and c. For the case $k = k_1$, the rotation of the phasor representing the multiplying factor e^{-ikx} is shown in Fig. 4.13b, and the result of multiplying that phasor by the phasor representing $\psi(x)$ is shown in Fig. 4.13c. As expected for the Fourier transform of the cosine function, the result is real, and its amplitude is given by the sum of the product phasors, several of which are shown in the figure.

So rotating phasors can be helpful in visualizing the process of Fourier transformation whether the function you're trying to analyze is real, imaginary, or complex.

To understand the connection of Fourier analysis to quantum wavefunctions, it's helpful to express these functions and the multiply-and-integrate process using Dirac notation. To do that, remember that the position-basis wavefunction $\psi(x)$ is the projection of a basis-independent state vector represented by ket $|\psi\rangle$ onto the position basis vector represented by ket $|x\rangle$:

$$\psi(x) = \langle x|\psi\rangle\,. \tag{4.18}$$

Note also that the plane-wave functions $\frac{1}{\sqrt{2\pi}}e^{ikx}$ are the wavenumber eigenfunctions represented in the position basis,[8] and may therefore be written as

[8] You can read more about position and wavenumber/momentum eigenfunctions in various basis systems in the final section of this chapter.

the projection of a basis-independent wavenumber vector represented by ket $|k\rangle$ onto the position basis vector $|x\rangle$:

$$\frac{1}{\sqrt{2\pi}} e^{ikx} = \langle x|k\rangle \,. \tag{4.19}$$

Rearranging the Fourier transform (Eq. 4.14) as

$$\phi(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \psi(x) e^{-ikx} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-ikx} \psi(x) dx$$

and then inserting $\langle x|k\rangle^*$ for $\frac{1}{\sqrt{2\pi}} e^{-ikx}$ and $\langle x|\psi\rangle$ for $\psi(x)$ gives

$$\begin{aligned} \phi(k) &= \int_{-\infty}^{\infty} \langle x|k\rangle^* \langle x|\psi\rangle \, dx = \int_{-\infty}^{\infty} \langle k|x\rangle \langle x|\psi\rangle \, dx \\ &= \langle k|\widehat{I}|\psi\rangle \,, \end{aligned} \tag{4.20}$$

in which $\widehat{I}$ represents the identity operator.

If you're wondering where the identity operator comes from in Eq. 4.20, note that $|x\rangle \langle x|$ is a projection operator (see Section 2.4); specifically, it's the operator that projects any vector onto the position basis vectors $|x\rangle$. And as described in Section 2.4, the sum (or integral in the continuous case) of a projection operator over all basis vectors gives the identity operator. So

$$\phi(k) = \langle k|\widehat{I}|\psi\rangle = \langle k|\psi\rangle \,. \tag{4.21}$$

In this representation, the result of the Fourier transform, the wavenumber spectrum $\phi(k)$, is expressed as the projection of the state represented by the ket $|\psi\rangle$ onto the wavenumber ket $|k\rangle$ through the inner product. In the position basis, the state $|\psi\rangle$ corresponds to the position-basis wavefunction $\psi(x)$, and the wavenumber ket $|k\rangle$ corresponds to the plane-wave sinusoidal basis functions $\frac{1}{\sqrt{2\pi}} e^{ikx}$.

Whether you think of the Fourier transform as multiplying a function by cosine and sine functions and integrating the products, or as multiplying and adding phasors, or as projecting an abstract state vector onto sinusoidal wavenumber basis functions, the bottom line is that Fourier analysis provides a means of determining the amount of each of the constituent sinusoidal waves that make up the function.

Fourier synthesis provides a complementary function: producing a wavefunction with desired characteristics by combining sinusoidal functions (in the form of e^{ikx}) in the proper proportions. This is useful, for example, in producing a "wave packet" with limited spatial extent.

Producing wave packets from monochromatic (single-wavenumber) plane waves is an important application of Fourier synthesis in quantum mechanics, because monochromatic plane-wave functions are not normalizable. That's because functions such as $Ae^{i(kx-\omega t)}$ extend to infinity in both the positive and negative x directions, which means the area under the square magnitude of such sinusoidal functions is infinite. As described earlier in this chapter, such non-normalizable functions are disqualified from serving as physically realizable quantum wavefunctions. But sinusoidal wavefunctions that are limited to a certain region of space are normalizable, and such wave-packet functions may be synthesized from monochromatic plane waves.

To do that, it's necessary to combine multiple plane waves in just the right proportions so that they add constructively in the desired region and destructively outside of that region. Those "right proportions" are provided by the continuous wavenumber function $\phi(k)$ inside the integral in the inverse Fourier transform:

$$\psi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \phi(k) e^{ikx} dk. \tag{4.15}$$

As described earlier, for each wavenumber k, the wavenumber spectrum $\phi(k)$ tells you the amount of the complex sinusoidal function e^{ikx} to add into the mix to synthesize $\psi(x)$.

You can think of the Fourier transform as a process that maps functions of one space or "domain" (such as position or time) onto another "domain" (such as wavenumber or frequency). Functions related by the Fourier transform ($\psi(x)$ and $\phi(k)$, for example), are sometimes called "Fourier-transform pairs," and the variables on which those functions depend (position x and wavenumber k in this case) are called "**conjugate variables**." Such variables always obey an uncertainty principle, which means that simultaneous precise knowledge of both variables is not possible. You can understand the reason for this by considering the Fourier-transform relationships illustrated in Figs. 4.14 – 4.18.

The position wavefunction $\psi(x)$ and wavenumber spectrum $\phi(k)$ of spatially limited quantum wave packets are an excellent example of Fourier-transform pairs and conjugate variables. To understand such wave packets, consider what happens if you add together the plane-wave functions e^{ikx}, each with amplitude of unity, over a range of wavenumbers Δk centered on the single wavenumber k_0. The wavenumber spectrum $\phi(k)$ for this case is shown in Fig. 4.14a.

The position wavefunction $\psi(x)$ corresponding to this wavenumber spectrum can be found using the inverse Fourier transform, so plug this $\phi(k)$ into Eq. 4.15:

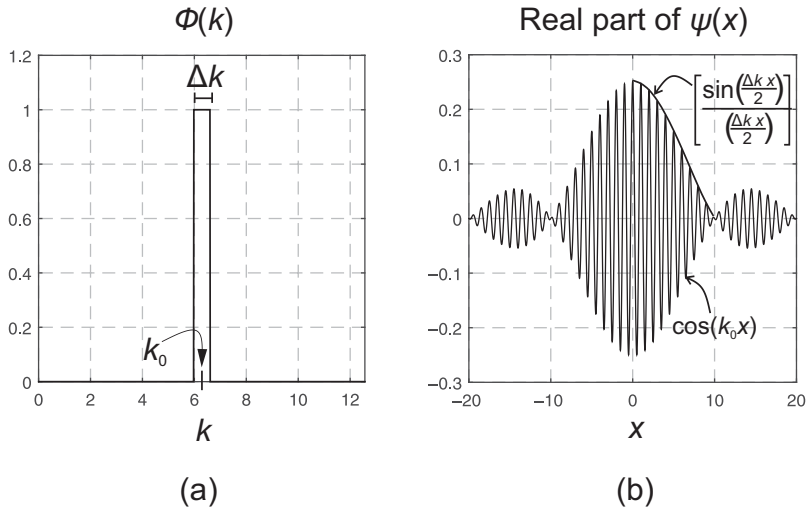


Figure 4.14 (a) wavenumber spectrum $\phi(k)$ producing (b) spatially limited wave packet $\psi(x)$.

$$\psi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \phi(k) e^{ikx} dk = \frac{1}{\sqrt{2\pi}} \int_{k_0 - \frac{\Delta k}{2}}^{k_0 + \frac{\Delta k}{2}} (1) e^{ikx} dk.$$

This integral is easily evaluated using $\int_a^b e^{cx} dx = \frac{1}{c} e^{cx} \Big|_a^b$, which makes the expression for $\psi(x)$

$$\begin{aligned} \psi(x) &= \frac{1}{\sqrt{2\pi}} \frac{1}{ix} e^{ikx} \Big|_{k_0 - \frac{\Delta k}{2}}^{k_0 + \frac{\Delta k}{2}} = \frac{-i}{\sqrt{2\pi}x} \left[e^{i(k_0 + \frac{\Delta k}{2})x} - e^{i(k_0 - \frac{\Delta k}{2})x} \right] \\ &= \frac{-i}{\sqrt{2\pi}x} e^{ik_0 x} \left[e^{i\frac{\Delta k}{2}x} - e^{-i\frac{\Delta k}{2}x} \right]. \end{aligned} \tag{4.22}$$

Now look at the term in square brackets, and recall that the inverse Euler relation for the sine function is

$$\sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i} \tag{4.23}$$

or

$$\left[e^{i\frac{\Delta k}{2}x} - e^{-i\frac{\Delta k}{2}x} \right] = 2i \sin\left(\frac{\Delta k}{2} x \right). \tag{4.24}$$

Inserting this into Eq. 4.22 gives

$$\psi(x) = \frac{-i}{\sqrt{2\pi}x} e^{ik_0 x}\left[e^{i\frac{\Delta k}{2}x} - e^{-i\frac{\Delta k}{2}x}\right] = \frac{-i}{\sqrt{2\pi}x} e^{ik_0 x}\left[2i\sin\left(\frac{\Delta k}{2}x\right)\right]$$
$$= \frac{2}{\sqrt{2\pi}x} e^{ik_0 x}\sin\left(\frac{\Delta k}{2}x\right). \tag{4.25}$$

The behavior of this expression as x changes can be understood more readily if you multiply both numerator and denominator by $\frac{\Delta k}{2}$ and do a bit of rearranging:

$$\psi(x) = \frac{(\frac{\Delta k}{2})2}{(\frac{\Delta k}{2})\sqrt{2\pi}x} e^{ik_0 x}\sin\left(\frac{\Delta k}{2}x\right)$$
$$= \frac{\Delta k}{\sqrt{2\pi}} e^{ik_0 x}\frac{\sin\left(\frac{\Delta k}{2}x\right)}{\left(\frac{\Delta k}{2}x\right)}. \tag{4.26}$$

Look carefully at the terms that vary with x. The first is $e^{ik_0 x}$, which has a real part equal to $\cos(k_0 x)$. So as x varies, this term oscillates between +1 and −1, with one cycle (2π of phase) in distance $\lambda_0 = 2\pi/k_0$. In Fig. 4.14b, λ_0 is taken as one distance unit, and you can see these rapid oscillations repeating at integer values of x.

Now think about the rightmost fraction in Eq. 4.26, which also varies with x. This term has the well-known form of $\frac{\sin(ax)}{ax}$ (which is why we multiplied numerator and denominator by $\Delta k/2$), called the **"sinc" function**. The sinc function has a large central region (sometimes called the "main lobe") and a series of smaller but significant maxima ("sidelobes") which decrease with distance from the central maximum. This function has its maximum at $x = 0$ (as you can verify using **L'Hôpital's rule**) and repeatedly crosses through zero between its lobes. The first zero-crossing of the sinc function occurs where the sine function in its numerator reaches zero, and the sine function hits zero where its argument equals π. So the first zero-crossing occurs when $(\Delta k/2)x = \pi$, or $x = 2\pi/\Delta k$.

There's an important point in that conclusion, because it's the sinc-function term in Eq. 4.26 that determines the spatial extent of the main lobe of the wave packet represented by $\psi(x)$. So to make a narrow wave packet, Δk must be large – that is, you need to include a large range of wavenumbers in the mix of plane waves that make up $\psi(x)$. And although the distance over which the wave packet decreases to a given value depends on the shape of the wavenumber spectrum $\phi(k)$, the conclusion that the width of the wave packet in x varies

inversely with the width of the wavenumber spectrum in k applies to spectra of all shapes, not just to the flat-topped wavenumber spectrum used in this example.

In the case shown in Fig. 4.14, Δk is taken to be 10% of k_0, so the first zero-crossing of $\psi(x)$ occurs at

$$x = \frac{2\pi}{\Delta k} = \frac{2\pi}{0.1k_0} = \frac{2\pi}{0.1\frac{2\pi}{\lambda_0}} = 10\lambda_0,$$

and since λ_0 is taken as one distance unit in this plot, the zero-crossing occurs at $x = 10$.

You can see the effect of increasing the width of the wavenumber spectrum in Fig. 4.15. In this case, Δk is increased to 50% of k_0, so the wavenumber spectrum $\phi(k)$ is five times wider than that of Fig. 4.14, and the wave packet envelope is narrower by that same factor (the first zero-crossing occurs at $x = 2$ distance units in this case).

Fig. 4.16 shows the effect of reducing the width of the wavenumber spectrum; in this case Δk is decreased to 5% of k_0, and you can see that the wave packet envelope is twice as wide of that of Fig. 4.14 (first zero crossing at $x = 20$).

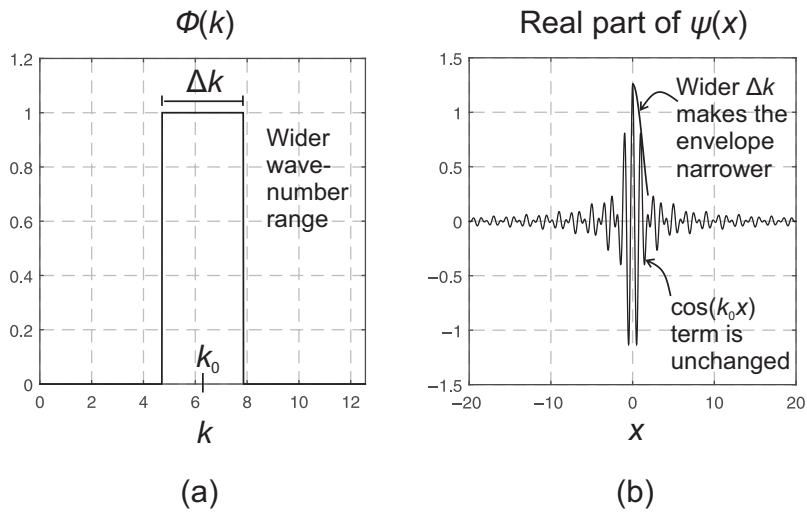


Figure 4.15 Summing a narrow range of wavenumbers, as shown in (a), reduces the width of the envelope of $\psi(x)$, as shown in (b).

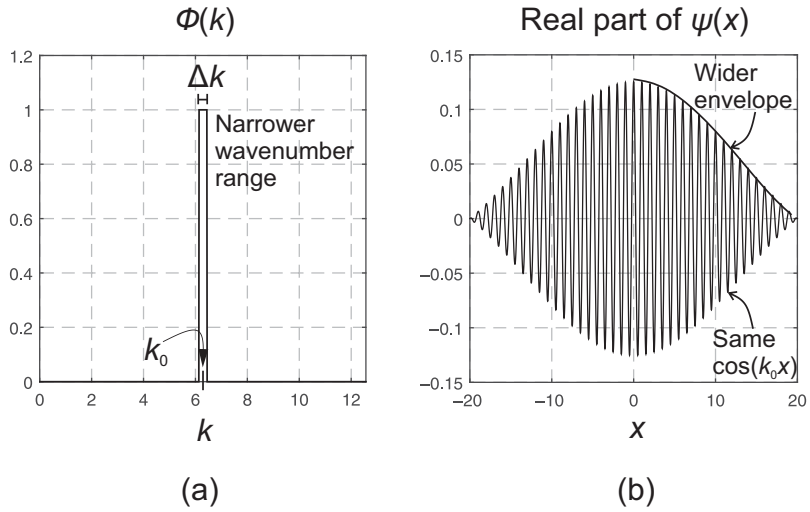


Figure 4.16 Summing a narrow range of wavenumbers, as shown in (a), increases the width of the envelope of $\psi(x)$, as shown in (b).

Even without knowing the inverse relationship between the widths of $\psi(x)$ and $\phi(k)$, you could have guessed what happens if you continue decreasing the width of the wavenumber spectrum, letting Δk approach zero. After all, if $\Delta k = 0$, then $\phi(k)$ consists of the single wavenumber k_0. And you know how a single-frequency (monochromatic) plane wave behaves: it extends to infinity in both the positive-x and negative-x directions. In other words, the "width" of $\psi(x)$ becomes infinite, as the first zero-crossing of the envelope never happens.

That behavior is illustrated in Fig. 4.17, for which the width Δk of the wavenumber spectrum has been reduced to 0.5% of k_0. Now the real part of $\psi(x)$ is essentially a pure cosine function, since the width of the sinc term in $\phi(k)$ is wider than the horizontal extent of the plot.

To see how this works mathematically, start with the definition of the inverse Fourier transform (Eq. 4.15) for a function of the wavenumber variable k' (the reason for the prime will become apparent shortly):

$$\psi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \phi(k') e^{ik'x} dk'. \tag{4.27}$$

Now plug this expression for $\psi(x)$ into the definition of the Fourier transform (Eq. 4.14):

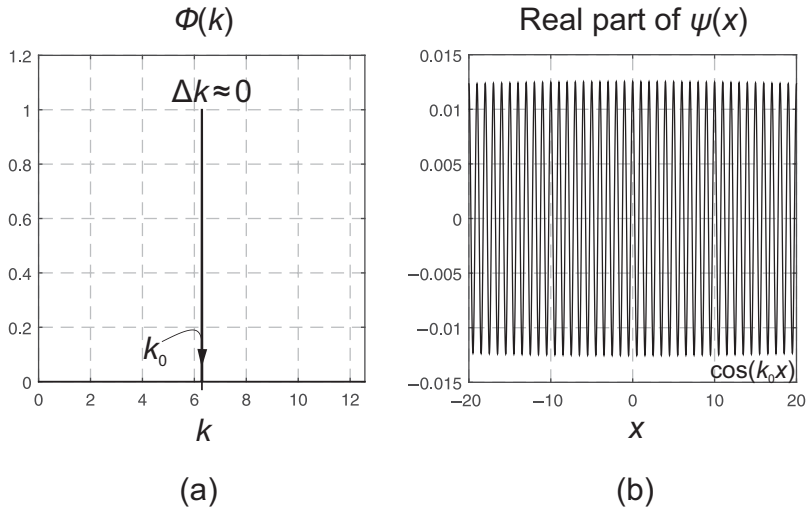


Figure 4.17 As the width of $\phi(k)$ approaches zero, as shown in (a), the width of the envelope of $\psi(x)$ approaches infinity, as shown in (b).

$$\begin{aligned}\phi(k) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\psi(x)e^{-ikx}dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\left[\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\phi(k')e^{ik'x}dk'\right]e^{-ikx}dx \end{aligned} \tag{4.28}$$

in which the prime on k' is included to help you discriminate between the wavenumbers over which the integral to form $\psi(x)$ is taken from the wavenumbers of the spectrum $\phi(k)$.

Eq. 4.28 looks a bit messy, but remember that you're free to move multiplicative terms inside or outside an integral as long as those terms don't depend on the variable over which the integral is being performed. Using that freedom to move the e^{-ikx} term inside the k' integral and combining the constants makes this

$$\phi(k) = \int_{-\infty}^{\infty}\left[\frac{1}{2\pi}\int_{-\infty}^{\infty}\phi(k')e^{i(k'-k)x}dk'\right]dx.$$

Now interchange the order of integration. Can you really do that? You can if the function under the integrals is continuous and the double integral is well-behaved. And in this case the limits of both integrals are $-\infty$ to ∞, which

means you don't have to fiddle with the limits when switching the order of integration. So

$$\phi(k) = \int_{-\infty}^{\infty} \phi(k') \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i(k'-k)x} dx \right] dk'. \tag{4.29}$$

Take a step back and consider the meaning of this expression. It's telling you that the value of the function $\phi(k)$ at every wavenumber k equals the integral of that same function ϕ multiplied by the term in square brackets and integrated over all wavenumbers k' from $-\infty$ to ∞. For that to be true, the term in square brackets must be performing a very unusual function: it must be "sifting" through the $\phi(k')$ function and pulling out the value $\phi(k)$. So in this case the integral ends up not summing anything; the function ϕ just takes on the value of $\phi(k)$ and walks straight out of the integral.

What magical function can perform such an operation? You've seen it before: it's the Dirac delta function, which can be defined as

$$\delta(x' - x) = \begin{cases} \infty, & \text{if } x' = x \\ 0, & \text{otherwise.} \end{cases} \tag{4.30}$$

A far more useful definition doesn't show what the Dirac delta function *is*; it shows what the Dirac delta function *does*:

$$\int_{-\infty}^{\infty} f(x')\delta(x' - x)dx' = f(x). \tag{4.31}$$

In other words, the Dirac delta function multiplied by a function inside an integral performs the exact sifting function needed in Eq. 4.29.

That means you can write Eq. 4.29 as

$$\phi(k) = \int_{-\infty}^{\infty} \phi(k') \left[\delta(k' - k) \right] dk', \tag{4.32}$$

and equating the terms in square brackets in Eqs. 4.32 and 4.29:

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i(k'-k)x} dx = \delta(k' - k). \tag{4.33}$$

This relationship is extremely useful when you're analyzing functions synthesized from combinations of sinusoids, as is another version that you can find by plugging the expression for $\phi(k)$ from Eq. 4.14 into the inverse Fourier transform (Eq. 4.15). That leads to[9]

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ik(x'-x)} dk = \delta(x' - x). \tag{4.34}$$

[9] If you need help getting this result, see the chapter-end problems and online solutions.

To see how these relationships can be useful, consider the process of taking the Fourier transform of a single-wavenumber (monochromatic) wavefunction $\psi(x) = e^{ik_0 x}$. Plugging this position wavefunction into Eq. 4.14 yields

$$\begin{aligned}\phi(k) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\psi(x)e^{-ikx}dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{ik_0 x}e^{-ikx}dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{i(k_0-k)x}dx = \sqrt{2\pi}\,\delta(k_0-k).\end{aligned} \tag{4.35}$$

So just as you'd expect from Fig. 4.17, the Fourier transform of the function $\psi(x) = e^{ik_0 x}$, which has finite spatial extent, is an infinitely narrow spike at wavenumber $k = k_0$.

If you're wondering about the amplitudes of the wavenumber spectrum $\phi(k)$ and position wavefunction $\psi(x)$ in Fig. 4.17, note that the maximum of $\phi(k)$ has been scaled to an amplitude of unity, and Eq. 4.26 shows that the amplitude of $\psi(x)$ is determined by the factor $\frac{\Delta k}{\sqrt{2\pi}}$. Since Δk has been set to 0.5% of k_0 and $k_0 = 2\pi$ in this case, the amplitude of $\psi(x)$ is $0.005(2\pi)/\sqrt{2\pi} = 0.0125$, as you can see in Fig. 4.17.

This is one extreme case: a spike with width approaching zero in the wavenumber domain is the Fourier transform of a sinusoidal function with width approaching infinity in the position domain. The other extreme is shown in Fig. 4.18; in this case the width Δk of the wavenumber spectrum has been increased so that $\phi(k)$ extends with constant amplitude from $k = 0$ to $2k_0$. As you can see in Fig. 4.18b, in this case the real part of $\psi(x)$ is close to a delta function $\delta(x)$ at position $x = 0$.

To see the mathematics of this case, here's what happens when you insert a narrow spike in position, $\psi(x) = \delta(x)$, into the Fourier transform (Eq. 4.14) to determine the corresponding wavenumber spectrum $\phi(k)$:

$$\phi(k) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\psi(x)e^{-ikx}dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\delta(x)e^{-ikx}dx. \tag{4.36}$$

But you know that the Dirac delta function under the integral sifts the function e^{-ikx}, so the only contribution comes from the function with $x = 0$:

$$\phi(k) = \frac{1}{\sqrt{2\pi}}e^0 = \frac{1}{\sqrt{2\pi}}. \tag{4.37}$$

This constant value means that $\phi(k)$ has uniform amplitude over all wavenumbers k, as shown in Fig. 4.18a. As in the previous figure, the amplitude of $\phi(k)$ has been scaled to a value of one, which is related to the maximum value of

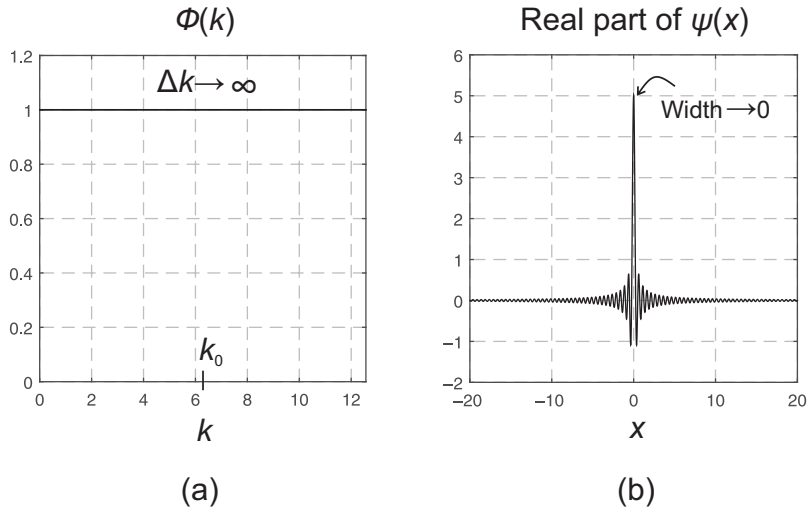


Figure 4.18 As the width of $\phi(k)$ approaches infinity, as shown in (a), the width of the envelope of $\psi(x)$ approaches zero, as shown in (b).

$\psi(x)$ by the factor $\frac{\Delta k}{\sqrt{2\pi}}$. That works out to 5.01 for this case, since $\Delta k = 2k_0$ and $k_0 = 2\pi$.

So as expected, a position function with extremely narrow width is a Fourier-transform pair with an extremely wide wavenumber function, just as a narrow wavenumber function pairs with a wide position function. This inverse relationship between the widths of functions of conjugate variables is the basis of the uncertainty principle, and in the next section, you'll see how the uncertainty principle applies to the conjugate variables of position and momentum.

4.5 Position and Momentum Wavefunctions and Operators

The presentation of wavefunction information in different spaces or domains, such as the position and wavenumber domains discussed in the previous section, is useful in many applications of physics and engineering. In quantum mechanics, the wavefunction representations you're likely to encounter include position and momentum, so this section is all about position and momentum wavefunctions, eigenfunctions, and operators – specifically, how to represent those functions and operators in both position space and momentum space.

You've already seen the connection between wavenumber (k) and momentum (p), which is provided by the de Broglie relation

$$p = \hbar k. \tag{3.4}$$

This means that the Fourier-transform relationship between functions of position and wavenumber also works between functions of position and momentum. Specifically, the momentum wavefunction $\tilde{\phi}(p)$ is the Fourier transform of the position wavefunction $\psi(x)$:

$$\tilde{\phi}(p) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} \psi(x) e^{-i\frac{p}{\hbar}x} dx, \tag{4.38}$$

in which $\tilde{\phi}(p)$ is a function of momentum (p)[10].

Additionally, the inverse Fourier transform of the momentum wavefunction $\tilde{\phi}(p)$ gives the position wavefunction $\psi(x)$:

$$\psi(x) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} \tilde{\phi}(p) e^{i\frac{p}{\hbar}x} dp. \tag{4.39}$$

Since $k = \frac{p}{\hbar}$, $dk = \frac{dp}{\hbar}$, and substituting $\frac{p}{\hbar}$ for k and $\frac{dp}{\hbar}$ for dk in Eq. 4.15 for the inverse Fourier transform yields

$$\psi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \tilde{\phi}(p) e^{i\frac{p}{\hbar}x} \frac{dp}{\hbar}, \tag{4.40}$$

which differs from Eq. 4.39 by a factor of $\frac{1}{\sqrt{\hbar}}$. In some texts (including this one), that factor is absorbed into the function $\tilde{\phi}$, but several popular quantum texts absorb $\frac{1}{\hbar}$ into $\tilde{\phi}$, omitting the factor of $\frac{1}{\sqrt{\hbar}}$ in the definitions of the Fourier transform and the inverse Fourier transform. In those texts, the factor in front of the integrals in Eqs. 4.38 and 4.39 is $\frac{1}{\sqrt{2\pi}}$.

Whichever convention you use for the constants, the relationship between position and momentum wavefunctions can help you understand one of the iconic laws of quantum mechanics. That law is the **Heisenberg Uncertainty principle**, which follows directly from the Fourier-transform relationship between position and momentum.

You can find an uncertainty principle for any Fourier-transform pair of conjugate wavefunctions, including the momentum-basis equivalent of the rectangular (flat-amplitude) wavenumber spectrum $\phi(k)$ and $\frac{\sin(ax)}{ax}$ position wavefunction discussed in the previous section. But it's also instructive to

[10] This notation is quite common in quantum textbooks; the tilde (˜) distinguishes the momentum wavefunction $\tilde{\phi}(p)$ from the wavenumber wavefunction $\phi(k)$.

consider a momentum wavefunction $\tilde{\phi}$ that doesn't produce an extended lobe structure in $\psi(x)$ such as that shown in Fig. 4.14b, since one of the goals of adding wavefunctions over a range of wavenumbers or momenta is to produce a spatially limited position wavefunction. So a position-space wavefunction that decreases smoothly toward zero amplitude without those extended lobes is desirable.

One way to accomplish that is to form a Gaussian wave packet. You may be wondering whether this means Gaussian in position space or Gaussian in momentum space, to which the answer is "both." To understand why that's true, start with the standard definition of a Gaussian function of position (x):

$$G(x) = Ae^{\frac{-(x-x_0)^2}{2\sigma_x^2}}, \tag{4.41}$$

in which A is the amplitude (maximum value) of $G(x)$, x_0 is the center location (x-value of the maximum), and σ_x is the standard deviation, which is half the width of the function between the points at which $G(x)$ is reduced to $\frac{1}{\sqrt{e}}$ (about 61%) of its maximum value.

Gaussian functions have several characteristics that make them instructive as quantum wavefunctions, including these two:

a) The square of a Gaussian is also a Gaussian, and

b) The Fourier transform of a Gaussian is also a Gaussian.

The first of these characteristics is useful because the probability density is related to the square of the wavefunction, and the second is useful because position-space and momentum-space wavefunctions are related by the Fourier transform.

You can see one of the benefits of the smooth shape of the Gaussian in Fig. 4.19. The Fourier-transform relationship between $\psi(x)$ and $\tilde{\phi}(p)$ means that smoothing the sharp corners of the rectangular momentum spectrum $\tilde{\phi}(p)$ significantly reduces the value of the magnitude of the position wavefunction in the region of the $\sin(ax)/ax$ lobe structure.

In position space, the phrase "Gaussian wave packet" means that the envelope of a sinusoidally varying function has a Gaussian shape. Such a packet can be formed by multiplying the Gaussian function $G(x)$ by the function $e^{i\frac{p_0}{\hbar}x}$ for a plane wave with momentum p_0:

$$\psi(x) = Ae^{\frac{-(x-x_0)^2}{2\sigma_x^2}} e^{i\frac{p_0}{\hbar}x}, \tag{4.42}$$

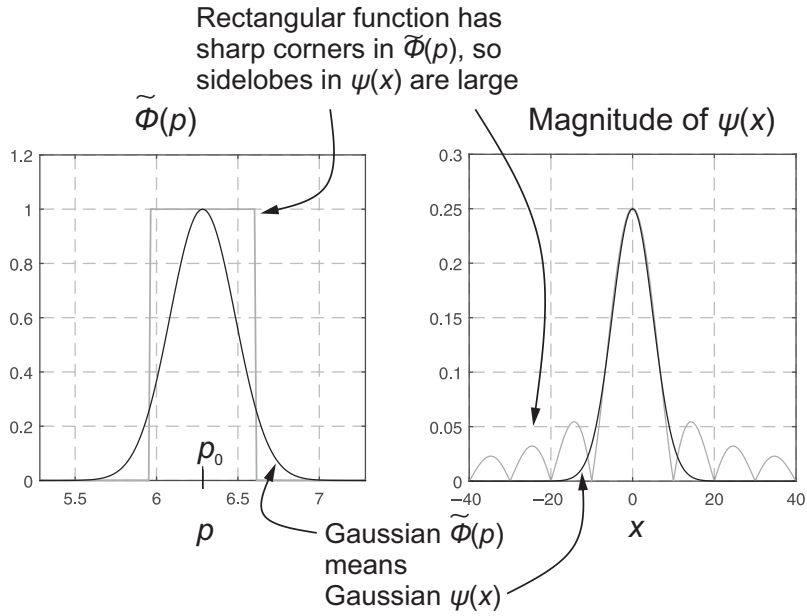


Figure 4.19 Improved spatial localization of position wavefunction $\psi(x)$ using Gaussian rather than rectangular function in momentum space.

in which the plane-wave amplitude has been absorbed into the constant A. When you're dealing with such a Gaussian wavefunction, it's important to realize that the quantity σ_x represents the standard deviation of the wavefunction $\psi(x)$, which is *not* the same as the standard deviation of the probability distribution that results from this wavefunction. That probability distribution is also a Gaussian, but with a different standard deviation, as you'll see later in this section.

When you're dealing with a quantum wavefunction, it's always a good idea to make sure that the wavefunction is normalized. Here's how that works for $\psi(x)$:

$$1 = \int_{-\infty}^{\infty} \psi^* \psi \, dx = \int_{-\infty}^{\infty} \left[A e^{\frac{-(x-x_0)^2}{2\sigma_x^2}} e^{i\frac{p_0}{\hbar}x} \right]^* \left[A e^{\frac{-(x-x_0)^2}{2\sigma_x^2}} e^{i\frac{p_0}{\hbar}x} \right] dx$$

$$= \int_{-\infty}^{\infty} |A|^2 \left[e^{\frac{-(x-x_0)^2}{\sigma_x^2}} \right] e^{\frac{(-p_0+p_0)x}{\hbar}} dx = |A|^2 \int_{-\infty}^{\infty} e^{\frac{-(x^2-2x_0x+x_0^2)}{\sigma_x^2}} dx.$$

This definite integral can be evaluated using

$$\int_{-\infty}^{\infty} e^{-(ax^2+bx+c)}dx = \sqrt{\frac{\pi}{a}}e^{\frac{b^2-4ac}{4a}}. \tag{4.43}$$

In this case $a = \frac{1}{\sigma_x^2}$, $b = \frac{-2x_0}{\sigma_x^2}$, and $c = \frac{x_0^2}{\sigma_x^2}$, so

$$1 = |A|^2\sqrt{\frac{\pi}{\frac{1}{\sigma_x^2}}}e^{\frac{(\frac{-2x_0}{\sigma_x^2})^2-4\frac{1}{\sigma_x^2}\frac{x_0^2}{\sigma_x^2}}{4\frac{1}{\sigma_x^2}}} = |A|^2\sqrt{\sigma_x^2\pi}e^{\frac{4x_0^2-4x_0^2}{4\sigma_x^2}} = |A|^2\sigma_x\sqrt{\pi}.$$

Solving for A yields

$$A = \frac{1}{(\sigma_x\sqrt{\pi})^{1/2}} \tag{4.44}$$

and the normalized position wavefunction is

$$\psi(x) = \frac{1}{(\sigma_x\sqrt{\pi})^{1/2}}e^{\frac{-(x-x_0)^2}{2\sigma_x^2}}e^{i\frac{p_0}{\hbar}x}. \tag{4.45}$$

To find the momentum wavefunction $\tilde{\phi}(p)$ corresponding to this normalized position wavefunction, take the Fourier transform of $\psi(x)$. To simplify the notation, you can take the origin of coordinates to be at x_0, so $x_0 = 0$. That makes the Fourier transform look like this:

$$\begin{aligned}\tilde{\phi}(p) &= \frac{1}{\sqrt{2\pi\hbar}}\int_{-\infty}^{\infty}\psi(x)e^{-i\frac{p}{\hbar}x}dx = \frac{1}{\sqrt{2\pi\hbar}}\int_{-\infty}^{\infty}\frac{1}{(\sigma_x\sqrt{\pi})^{1/2}}e^{\frac{-x^2}{2\sigma_x^2}}e^{-i\frac{p-p_0}{\hbar}x}dx \\ &= \frac{1}{\sqrt{2\pi\hbar}}\frac{1}{(\sigma_x\sqrt{\pi})^{1/2}}\int_{-\infty}^{\infty}e^{\frac{-x^2}{2\sigma_x^2}-i\frac{p-p_0}{\hbar}x}dx.\end{aligned}$$

Using the same definite integral given earlier in this section with $a = \frac{1}{2\sigma_x^2}$, $b = -i\frac{p-p_0}{\hbar}$, and $c = 0$ gives

$$\begin{aligned}\tilde{\phi}(p) &= \frac{1}{\sqrt{2\pi\hbar}}\frac{1}{(\sigma_x\sqrt{\pi})^{1/2}}\sqrt{\frac{\pi}{a}}e^{\frac{b^2-4ac}{4a}} = \frac{1}{\sqrt{2\pi\hbar}}\frac{\sqrt{2\pi\sigma_x^2}}{(\sigma_x\sqrt{\pi})^{1/2}}e^{\frac{-(p-p_0)^2\sigma_x^2}{2\hbar^2}} \\ &= \left(\frac{\sigma_x^2}{\pi\hbar^2}\right)^{\frac{1}{4}}e^{\frac{-(p-p_0)^2\sigma_x^2}{2\hbar^2}}.\end{aligned}$$

This is also a Gaussian, since it can be written

$$\tilde{\phi}(p) = \left(\frac{\sigma_x^2}{\pi\hbar^2}\right)^{\frac{1}{4}}e^{\frac{-(p-p_0)^2}{2\sigma_p^2}}, \tag{4.46}$$

in which the **standard deviation** of the momentum wavefunction is given by $\sigma_p = \frac{\hbar}{\sigma_x}$.

Multiplying the standard deviations of these Gaussian position and momentum wavefunctions gives

$$\sigma_x \sigma_p = \sigma_x \left(\frac{\hbar}{\sigma_x} \right) = \hbar. \tag{4.47}$$

It takes just one more step to get to the Heisenberg Uncertainty principle. To make that step, note that the "uncertainty" in the Heisenberg Uncertainty principle is defined with respect to the width of the probability distribution, which is narrower than the width of the Gaussian wavefunction $\psi(x)$.

To determine the relationship between these two different widths, remember that the probability density is proportional to $\psi^* \psi$. That means that the width Δx of the probability distribution can be found from

$$e^{-\frac{x^2}{2(\Delta x)^2}} = \left(e^{-\frac{x^2}{2\sigma_x^2}} \right)^* \left(e^{-\frac{x^2}{2\sigma_x^2}} \right) = e^{-\frac{x^2}{\sigma_x^2}}. \tag{4.48}$$

So $2(\Delta x)^2 = \sigma_x^2$, or $\sigma_x = \sqrt{2}\Delta x$. The same argument applies to the momentum-space wavefunction $\tilde{\phi}(p)$, so it's also true that $\sigma_p = \sqrt{2}\Delta p$, in which Δp represents the width of the probability distribution in momentum space.

This is the reason that many instructors and authors define the exponential term in the position wavefunction $\psi(x)$ as $e^{\frac{-(x-x_0)^2}{4\sigma_x^2}}$. In that case, the σ_x that they write in the exponential of $\psi(x)$ is the standard deviation of the probability distribution rather than the standard deviation of the wavefunction.

Writing Eq. 4.47 in terms of the widths of the probability distributions in position (Δx) and momentum (Δp) gives

$$\sigma_x \sigma_p = (\sqrt{2}\Delta x)(\sqrt{2}\Delta p) = \hbar \tag{4.49}$$

or

$$\Delta x \Delta p = \frac{\hbar}{2}. \tag{4.50}$$

This is the uncertainty relation for Gaussian wavefunctions. For any other functions, the product of the standard deviations gives a value greater than this, so the general uncertainty relation between conjugate variables such as

position and momentum (or any other two variables related by the Fourier transform) is

$$\Delta x \Delta p \geq \frac{\hbar}{2}. \tag{4.51}$$

This is the usual form of the Heisenberg Uncertainty principle. It says that for this pair of conjugate or "incompatible" observables, there is a fundamental limit to the precision with which both may be known. So precise knowledge of position (small Δx) is incompatible with precise knowledge of momentum (small Δp), since the product of their probability-distribution uncertainties ($\Delta x \Delta p$) must be equal to or larger than half of the reduced Planck constant $\hbar$.

Another important aspect of incompatible observables concerns the operators associated with those observables. Specifically, the operators of incompatible observables do not commute, which means that the order in which those operators are applied matters. To see why that's true for the position and momentum operators, it helps to have a good understanding of the form and behavior of these operators in both position and momentum space.

Students learning quantum mechanics often express confusion about quantum operators and their eigenfunctions, and that confusion is frequently embodied in questions such as:

- Why is the result of operating on a position wavefunction with the position operator $\widehat{X}$ equal to that wavefunction multiplied by x?
- Why are position eigenfunctions given by delta functions $\delta(x - x_0)$ in position space?
- Why is the result of operating on a momentum wavefunction with the momentum operator $\widehat{P}$ equal to the spatial derivative of that function multiplied by $-i\hbar$?
- Why are momentum eigenfunctions given by $\frac{1}{\sqrt{2\pi\hbar}} e^{i\frac{p}{\hbar}x}$ in position space?

To answer these questions, start by considering how an operator and its eigenfunctions are related to the expectation value of the observable associated with the operator. As explained in Section 2.5, the expectation value for a continuous observable such as position x is given by

$$\langle x \rangle = \int_{-\infty}^{\infty} xP(x)dx, \tag{4.52}$$

in which $P(x)$ represents the probability density as a function of position x.

For normalized quantum wavefunctions $\psi(x)$, the probability density is given by the square magnitude of the wavefunction $|\psi(x)|^2 = \psi(x)^*\psi(x)$, so the expectation value may be written as

$$\langle x \rangle = \int_{-\infty}^{\infty} x|\psi(x)|^2 dx = \int_{-\infty}^{\infty} [\psi(x)]^* x[\psi(x)]dx. \tag{4.53}$$

Compare this to the expression from Section 2.5 for the expectation value of an observable x associated with operator $\widehat{X}$ using the inner product:

$$\langle x \rangle = \langle \psi | \widehat{X} | \psi \rangle = \int_{-\infty}^{\infty} [\psi(x)]^* \widehat{X}[\psi(x)]dx. \tag{2.60}$$

For these expressions to be equal, the result of the operator $\widehat{X}$ acting on the wavefunction $\psi(x)$ must be to multiply $\psi(x)$ by x. And why does it do that? Because an operator's job is to pull out the eigenvalues (that is, the possible results of observations) from the eigenfunctions of that operator, as described in Section 4.2. In the case of the position observations, the possible results of measurement are every position x, so that's what the position operator $\widehat{X}$ pulls out from its eigenfunctions.

And what are those eigenfunctions of the $\widehat{X}$ operator? To answer that question, consider how those eigenfunctions behave. The eigenvalue equation for the position operator acting on the first of the eigenfunctions ($\psi_1(x)$) is

$$\widehat{X}\psi_1(x) = x_1\psi_1(x), \tag{4.54}$$

in which x_1 represents the eigenvalue associated with eigenfunction ψ_1. But since the action of the position operator is to multiply the function upon which it's operating by x, it must also be true that

$$\widehat{X}\psi_1(x) = x\psi_1(x). \tag{4.55}$$

Setting the right sides of Eqs. 4.54 and 4.55 equal to one another gives

$$x\psi_1(x) = x_1\psi_1(x). \tag{4.56}$$

Think about what this equation means: the variable x times the first eigenfunction ψ_1 is equal to the single eigenvalue x_1 times that same function. Since x varies over all possible positions while x_1 represents only a single position, how can this statement be true?

The answer is that the eigenfunction $\psi_1(x)$ must be zero everywhere except at the single location $x = x_1$. That way, when the value of x is not equal to x_1, both sides of Eq. 4.56 are zero, and the equation is true. And when $x = x_1$, this equation says $x_1\psi_1(x) = x_1\psi_1(x)$, which is also true.

So what function is zero for all values of x except when $x = x_1$? The Dirac delta function $\delta(x - x_1)$. And for the second eigenfunction $\psi_2(x)$ with eigenvalue x_2, the delta function $\delta(x - x_2)$ does the trick, as does $\delta(x - x_3)$ for $\psi_3(x)$, and so forth.

Thus the eigenfunctions of the position operator $\widehat{X}$ are an infinite set of Dirac delta functions $\delta(x - x')$, each with its own eigenvalue, and those eigenvalues (represented by x') cover the entire range of positions from $-\infty$ to $+\infty$.

You can bring this same analysis to bear on momentum operators and eigenfunctions, which behave in momentum space in the same way as position operators and eigenfunctions behave in position space.

That means that you can find the expectation value of momentum using the integral of the possible outcomes p times the probability density

$$\langle p \rangle = \int_{-\infty}^{\infty} p|\tilde{\phi}(p)|^2 dp = \int_{-\infty}^{\infty} [\tilde{\phi}(p)]^* p[\tilde{\phi}(p)]dp. \tag{4.57}$$

You can also write the momentum expectation value using the inner product, with the momentum-space representation of the momentum operator $\widehat{P}_p$ acting on the momentum-basis wavefunction $\tilde{\phi}(p)$:

$$\langle p \rangle = \left\langle \tilde{\phi} \middle| \widehat{P}_p \middle| \tilde{\phi} \right\rangle = \int_{-\infty}^{\infty} [\tilde{\phi}(p)]^* \widehat{P}_p [\tilde{\phi}(p)]dp. \tag{4.58}$$

In the notation $\widehat{P}_p$, the uppercase P wearing a hat tells you that this is the momentum operator, and the lowercase p in the subscript tells you that this is the momentum-basis version of the operator.

And just as in the case of the position operator, the action of the momentum operator is to multiply the function upon which it's operating by p. Hence for the eigenfunction $\tilde{\phi}_1$ with eigenvalue p_1

$$\widehat{P}_p \tilde{\phi}_1(p) = p\tilde{\phi}_1(p) = p_1 \tilde{\phi}_1(p). \tag{4.59}$$

For this equation to be true, the eigenfunction $\tilde{\phi}_1(p)$ must be zero everywhere except at the single location $p = p_1$. Thus the eigenfunctions of the momentum operator $\widehat{P}_p$ in momentum space are an infinite set of Dirac delta functions $\delta(p - p')$, each with its own eigenvalue, and those eigenvalues (represented by p') cover the entire range of momenta.

Here are the important points: in any operator's own space, the action of that operator on each of its eigenfunctions is to multiply that eigenfunction by the observable corresponding to the operator. And in the operator's own space, those eigenfunctions are Dirac delta functions.

This explains the form and behavior of operators and their eigenfunctions in their own space. But it's often useful to apply an operator to functions that reside in other spaces – for example, applying the momentum operator $\widehat{P}$ to position wavefunctions $\psi(x)$.

Why might you want to do that? Perhaps you have the position-basis wavefunction and you wish to find the expectation value of momentum. You can do that using the position-space representation of the momentum operator $\widehat{P}_x$ operating on the position-basis wavefunction $\psi(x)$

$$\langle p \rangle = \int_{-\infty}^{\infty} [\psi(x)]^* \widehat{P}_x [\psi(x)] dx, \tag{4.60}$$

in which lowercase x in the subscript of $\widehat{P}_x$ tells you that this is the position-basis version of the momentum operator $\widehat{P}$. This equation is the position-space equivalent to the momentum-space relation for the expectation value of p shown in Eq. 4.58.

So what is the form of the momentum operator $\widehat{P}$ in position space? One way to discover that is to begin with the eigenfunctions of that operator. Since you know that in momentum space the eigenfunctions of the momentum operator are the Dirac delta functions $\delta(p-p')$, you can use the inverse Fourier transform to find the position-space momentum eigenfunctions:

$$\begin{aligned}\psi(x) &= \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} \tilde{\phi}(p) e^{i\frac{p}{\hbar}x} dp = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} \delta(p-p') e^{i\frac{p}{\hbar}x} dp \\ &= \frac{1}{\sqrt{2\pi\hbar}} e^{i\frac{p'}{\hbar}x},\end{aligned}$$

in which p' is the continuous variable representing all possible values of momentum. Naming that variable p instead of p' makes the position representation of the momentum eigenfunctions

$$\psi_p(x) = \frac{1}{\sqrt{2\pi\hbar}} e^{i\frac{p}{\hbar}x}, \tag{4.61}$$

in which the subscript "p" is a reminder that these are the momentum eigenfunctions represented in the position basis. You can use this position-space representation of the momentum eigenfunctions to find the position-space representation $\widehat{P}_x$ of the momentum operator $\widehat{P}$. To do that, remember that the action of the momentum operator on its eigenfunctions is to multiply those eigenfunctions by p:

$$\widehat{P}_x \psi_p(x) = p\psi_p(x). \tag{4.62}$$

Plugging in the position-space representations of the momentum eigenfunctions for $\psi_p(x)$ makes this

$$\widehat{P}_x \left[\frac{1}{\sqrt{2\pi\hbar}} e^{i\frac{p}{\hbar}x} \right] = p \left[\frac{1}{\sqrt{2\pi\hbar}} e^{i\frac{p}{\hbar}x} \right]. \tag{4.63}$$

The p that the operator must pull out of the eigenfunction is in the exponential, which suggests that a spatial derivative may be useful:

$$\frac{\partial}{\partial x}\left[\frac{1}{\sqrt{2\pi\hbar}}e^{i\frac{p}{\hbar}x}\right] = i\frac{p}{\hbar}\left[\frac{1}{\sqrt{2\pi\hbar}}e^{i\frac{p}{\hbar}x}\right].$$

So the spatial derivative does bring out a factor of p, but two constants come along with it. You can deal with those by multiplying both sides by $\frac{\hbar}{i}$:

$$\frac{\hbar}{i}\frac{\partial}{\partial x}\left[\frac{1}{\sqrt{2\pi\hbar}}e^{i\frac{p}{\hbar}x}\right] = \frac{\hbar}{i}\left(i\frac{p}{\hbar}\right)\left[\frac{1}{\sqrt{2\pi\hbar}}e^{i\frac{p}{\hbar}x}\right] = p\left[\frac{1}{\sqrt{2\pi\hbar}}e^{i\frac{p}{\hbar}x}\right]$$

exactly as needed. So position-space representation of the momentum operator $\widehat{P}$ is

$$\widehat{P}_x = \frac{\hbar}{i}\frac{\partial}{\partial x} = -i\hbar\frac{\partial}{\partial x}. \tag{4.64}$$

This is the form of the momentum operator $\widehat{P}$ in position space, and you can use $\widehat{P}_x$ to operate on position-basis wavefunctions $\psi(x)$.

The same approach can be used to determine the form of the position operator $\widehat{X}$ and its eigenfunctions in momentum space. This leads to position eigenfunctions in momentum space:

$$\tilde{\phi}_x(p) = \frac{1}{\sqrt{2\pi\hbar}}e^{-i\frac{p}{\hbar}x} \tag{4.65}$$

and the momentum-space representation $\widehat{X}_p$ of the position operator $\widehat{X}$:

$$\widehat{X}_p = i\hbar\frac{\partial}{\partial p}. \tag{4.66}$$

If you need help getting to these expressions, check out the problems at the end of this chapter and the online solutions.

Given these position-basis representations of the position and momentum operators, you can determine an important quantity in quantum mechanics. That quantity is the commutator $[\widehat{X}, \widehat{P}]$:

$$[\widehat{X}, \widehat{P}] = \widehat{X}\widehat{P} - \widehat{P}\widehat{X} = x(-i\hbar)\frac{d}{dx} - (-i\hbar)\frac{d}{dx}x. \tag{4.67}$$

Trying to analyze this expression in this form leads many students astray. To correctly determine the commutator, you should *always* provide a function

on which the operators can operate, like this:

$$[\widehat{X},\widehat{P}]\psi = (\widehat{X}\widehat{P} - \widehat{P}\widehat{X})\psi = \left[x(-i\hbar)\frac{d}{dx} - (-i\hbar)\frac{d}{dx}x\right]\psi$$
$$= x(-i\hbar)\frac{d\psi}{dx} - (-i\hbar)\frac{d(x\psi)}{dx}.$$

You can see the reason for inserting the function ψ in the last term – it reminds you that the spatial derivative d/dx must be applied not only to x, but to the product $x\psi$:

$$[\widehat{X},\widehat{P}]\psi = x(-i\hbar)\frac{d\psi}{dx} - (-i\hbar)\frac{d(x\psi)}{dx} = (-i\hbar)x\frac{d\psi}{dx} - (-i\hbar)\frac{d(x)}{dx}\psi - (-i\hbar)\frac{d\psi}{dx}x$$
$$= (-i\hbar)x\frac{d\psi}{dx} - (-i\hbar)(1)\psi - (-i\hbar)\frac{d\psi}{dx}x$$
$$= i\hbar\psi.$$

Now that the wavefunction ψ has done its job of helping you take all the required derivatives, you can remove it and write the commutator of the position and momentum operators as

$$[\widehat{X},\widehat{P}] = i\hbar. \tag{4.68}$$

Using the momentum-space representation of the operators $\widehat{X}$ and $\widehat{P}$ leads to the same result, as you can see in the chapter-end problems and online solutions.

This nonzero value of the commutator $[\widehat{X},\widehat{P}]$ (called the "canonical commutation relation") has extremely important implications, since it shows that the order in which certain operators are applied matters. Operators such as $\widehat{X}$ and $\widehat{P}$ are "non-commuting," which means they don't share the same eigenfunctions. Remember that the process of making a position measurement of a quantum observable for a particle or system in a given state causes the wavefunction to collapse to an eigenfunction of the position operator. But since the position and momentum operators don't commute, that position eigenfunction is not an eigenfunction of momentum. So if you then make a momentum measurement, the wavefunction (not being in a momentum eigenfunction) collapses to a momentum eigenfunction. That means the system is now in a different state, so your position measurement is no longer relevant. This is the essence of quantum indeterminacy.

In the next chapter, the quantum wavefunctions for three specific potentials are derived and explored. Before getting to that, here are some problems to help you apply the concepts discussed in this chapter.

4.6 Problems

1. Determine whether each of the following functions meets the requirements of a quantum wavefunction:

a) $f(x) = \frac{1}{(x-x_0)^2}$ over the range $x = -\infty$ to $+\infty$.
b) $g(x) = \sin(kx)$ over the range $x = -\pi$ to π (k is finite).
c) $h(x) = \sin^{-1}(x)$ over the range $x = -1$ to 1.
d) $\psi(x) = Ae^{ikx}$ (constant A) over the range $x = -\infty$ to $+\infty$.

2. Use the sifting property of the Dirac delta function to evaluate these integrals:

a) $\int_{-\infty}^{\infty} Ax^2 e^{ikx}\delta(x - x_0)dx$.
b) $\int_{-\infty}^{\infty} \cos(kx)\delta(k' - k)dk$.
c) $\int_{-2}^{3} \sqrt{x}\delta(x + 3)dx$.

3. Show that the Fourier-transform relationship between the position-space and momentum-space representations of the state represented by ket $|\psi\rangle$ can be written as

$$\tilde{\phi}(p) = \langle p|\psi\rangle = \int_{-\infty}^{\infty} \langle p|x\rangle \langle x|\psi\rangle \, dx$$

and

$$\psi(x) = \langle x|\psi\rangle = \int_{-\infty}^{\infty} \langle x|p\rangle \langle p|\psi\rangle \, dp.$$

4. Use Eq. 4.53 to find the expectation value $\langle x\rangle$ for a particle with position-basis wavefunction $\psi(x) = \sqrt{\frac{2}{a}} \sin\left(\frac{2\pi x}{a}\right)$ over the range $x = 0$ to $x = a$ and zero elsewhere.

5. Show that in two regions of piecewise-constant potential the amplitude ratio of the wavefunctions (such as $\psi(x)$ given by Eq. 4.10) on opposite sides of the boundary between the regions is inversely proportional to the wavenumber ratio (assume $E > V$ on both sides of the boundary).

6. (a) Show that the expressions $A_1 \cos(kx) + B_1 \sin(kx)$ and $A_2 \sin(kx + \phi)$ are equivalent to the expression $Ae^{ikx} + Be^{-ikx}$, and find the relationship between the coefficients of these expressions.

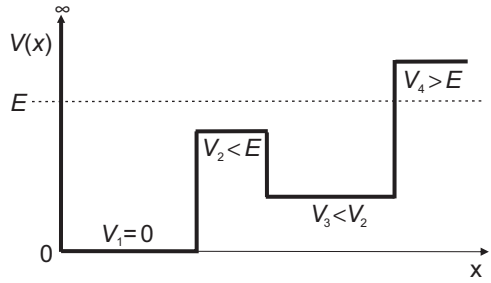
(b) Use L'Hôpital's rule to find the value of the function $\frac{\sin\left(\frac{\Delta k}{2}x\right)}{\frac{\Delta k}{2}x}$ at $x = 0$.

7. Show that plugging the expression for $\phi(k)$ from the Fourier transform (Eq. 4.14) into the inverse Fourier transform (Eq. 4.15) leads to the Dirac delta-function expression given in Eq. 4.34.

8. Derive the momentum-space representation of the position eigenfunctions $\tilde{\phi}(p)$ (Eq. 4.65) and the position operator $\widehat{X}$ (Eq. 4.66).

9. Use the momentum-space representations of the position and momentum operators to find the commutator $[\widehat{X}, \widehat{P}]$.

10. Given the piecewise-constant potential $V(x)$ shown in the figure, sketch the wavefunction $\psi(x)$ for a particle with energy E in each region.



5
Solutions for Specific Potentials

The conclusions reached in the previous chapter concerning quantum wavefunctions and their general behavior are based on the form of the Schrödinger equation, which relates the changes in a particle or system's wavefunction over space and time to the energy of that particle or system. Those conclusions tell you a great deal about how matter and energy behave at the quantum level, but if you want to make specific predictions about the outcome of measurements of observables such as position, momentum, and energy, you need to know the exact form of the potential energy in the region of interest. In this chapter, you'll see how to apply the concepts and mathematical formalism described in earlier chapters to quantum systems with three specific potentials: the infinite rectangular well, the finite rectangular well, and the harmonic oscillator.

Of course, you can find a great deal more information about each of these topics in comprehensive quantum texts and online. So the purpose of this chapter is not to provide one more telling of the same story; instead, these example potentials are meant to show why techniques such as taking the inner product between functions, finding eigenfunctions and eigenvalues of an operator, and using the Fourier transform between position and momentum space are so important in solving problems in quantum mechanics. As in previous chapters, the focus will be on the relationship between the mathematics of the solutions to the Schrödinger equation and the physical meaning of those solutions. And although we live in a universe with (at least) three spatial dimensions in which the potential energy $V(\vec{r}, t)$ may vary over time as well as space, most of the essential physics of quantum potential wells can be understood by examining the one-dimensional case with time-independent potential energy. So in this chapter, the Schrödinger equation is written with position represented by x and potential energy by $V(x)$.

5.1 Infinite Rectangular Potential Well

The infinite rectangular well is a potential configuration in which a quantum particle is confined to a specified region of space (called the "potential well") by infinitely strong forces at the edges of that region. Within the well, no force acts on the particle. Of course, this configuration is not physically realizable, since infinite forces do not occur in nature. But as you'll see in this section, the infinite rectangular potential well has several features that make this a highly instructive configuration.

Recall from classical mechanics that force $\vec{F}$ is related to potential energy V by the equation $\vec{F} = -\vec{\nabla}V$, in which $\vec{\nabla}$ represents the gradient differential operator (as described in Section 3.4, $\vec{\nabla} \equiv \hat{x}\frac{\partial}{\partial x} + \hat{y}\frac{\partial}{\partial y} + \hat{z}\frac{\partial}{\partial z}$ in 3-D Cartesian coordinates). So at the edges of the infinite rectangular well, infinite force means that the change in potential energy with distance must be infinite, while inside the well, zero force means that the potential energy must be constant. Since you're free to define the reference level for potential energy at any location, it's convenient to take the potential energy to be zero inside the well.

For a one-dimensional infinite rectangular well extending from $x = 0$ to $x = a$, the potential energy may be written

$$V(x) = \begin{cases} \infty, & \text{if } x < 0 \text{ or } x > a \\ 0, & 0 \le x \le a \end{cases} \tag{5.1}$$

and you can see the potential energy and forces in the region of such a one-dimensional infinite potential well[1] in Fig. 5.1.

Notice that as you move along the x-axis from left to right, the potential energy drops from infinity (in the region $x < 0$) to zero at the left wall (where $x = 0$). This means that $\frac{\partial V}{\partial x}$ equals negative infinity at $x = 0$, so the force (which in the 1-D case is *negative* $\frac{\partial V}{\partial x}$) has infinite magnitude and points in the positive-x direction. Moving along x within the well, $\frac{\partial V}{\partial x} = 0$, but at the right wall (where $x = a$) the potential energy increases from zero to infinity. This means that the change in potential energy at $x = a$ is infinitely positive, and that makes $-\frac{\partial V}{\partial x}$ infinitely negative at that location. So at the right wall, the force is again infinitely strong but pointing in the negative-x direction. Hence any particle within the well is "trapped" by infinitely strong inward-pointing forces at both walls.

[1] This configuration is sometimes called an infinite "square well," although the well is infinitely deep and therefore not square – the word "square" probably comes from the flat "bottom" and vertical "walls" of the well and the 90° angles at the base of each wall.

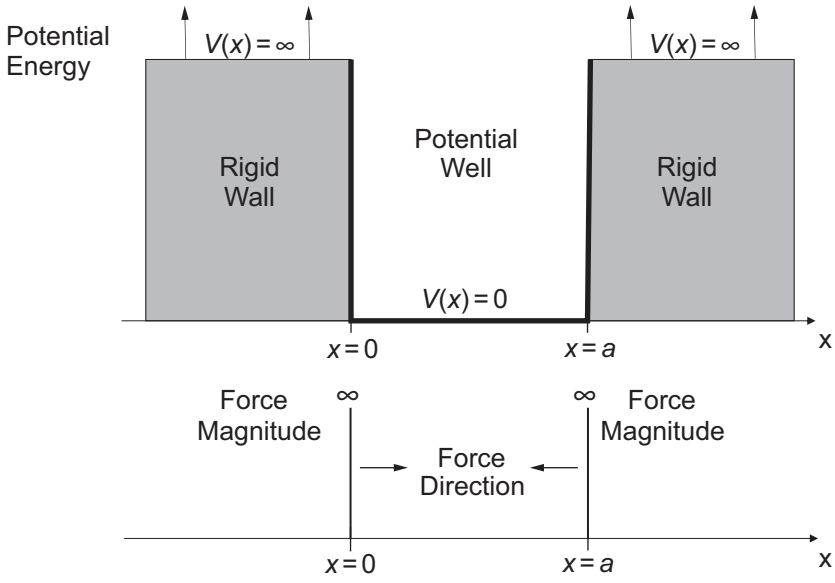


Figure 5.1 Infinite rectangular potential well.

Two unrealistic aspects of this configuration are the infinite potential energy outside the well and the infinite slope of the potential energy at each wall. The Schrödinger equation cannot be solved at the locations at which the potential energy makes an infinite jump, but meaningful results can still be obtained by finding the wavefunction solutions to the Schrödinger equation within and outside the well, and then joining those wavefunctions together at the edges of the well.

The infinite rectangular well is a good first example because it can be used to demonstrate useful techniques for solving both the time-dependent and the time-independent Schrödinger equation (TISE), and for understanding the behavior of quantum wavefunctions in position space and momentum space. Additionally, you can apply these same techniques to more realistic configurations that involve particles confined to a certain region of space by large (but finite) forces, such as an electron trapped by a strong electrostatic field.

To determine the behavior of particles in an infinite rectangular well, the first order of business is to find the possible wavefunctions of those particles. In this context, "possible" wavefunctions are those that are solutions of the Schrödinger equation and that satisfy the boundary conditions of the infinite rectangular well. And although the infinite-slope walls of such a well mean

that the slope of the wavefunction is not continuous at the boundaries, you can still solve the Schrödinger equation within the well (where the potential energy is zero) and enforce the boundary condition of continuous amplitude across the boundaries of the well.

As described in Section 3.3, it's often possible to determine the wavefunction solutions $\Psi(x,t)$ using separation of variables, and that's true in this case. So just as in Section 4.3, you can write the wavefunction $\Psi(x,t)$ as the product of a spatial function $\psi(x)$ and a temporal function $T(t)$: $\Psi(x,t) = \psi(x)T(t)$. This leads to the time-independent Schrödinger equation

$$-\frac{\hbar^2}{2m}\frac{d^2[\psi(x)]}{dx^2} + V[\psi(x)] = E[\psi(x)], \tag{3.40}$$

in which E is the separation constant connecting the separated temporal and spatial differential equations. The solutions to the TISE are the eigenfunctions of the Hamiltonian (total-energy) operator, and the eigenvalues associated with those eigenfunctions are the possible outcomes of energy measurements of a particle trapped in an infinite rectangular well.

Recall also from Section 4.3 that in regions in which $E > V$ it's convenient to write this as

$$\frac{d^2[\psi(x)]}{dx^2} = -\frac{2m}{\hbar^2}(E-V)\psi(x) = -k^2\psi(x), \tag{4.8}$$

in which the constant k is a wavenumber given by

$$k \equiv \sqrt{\frac{2m}{\hbar^2}(E-V)}. \tag{4.9}$$

The exponential form of the general solution to Eq. 4.8 is

$$\psi(x) = Ae^{ikx} + Be^{-ikx}, \tag{4.10}$$

in which A and B are constants to be determined by the boundary conditions.

Inside the infinite rectangular well, $V = 0$, so any positive value of E is greater than V, and the wavenumber k is

$$k = \sqrt{\frac{2m}{\hbar^2}E}, \tag{5.2}$$

which means that inside the well the wavefunction $\psi(x)$ oscillates with wavenumber proportional to the square root of the energy E.

The case in which $V > E$ is also considered in Section 4.3; in that case the TISE may be written as

$$\frac{d^2[\psi(x)]}{dx^2} = -\frac{2m}{\hbar^2}(E - V)\psi(x) = +\kappa^2\psi(x), \tag{4.11}$$

in which the constant κ is given by

$$\kappa \equiv \sqrt{\frac{2m}{\hbar^2}(V - E)}. \tag{4.12}$$

The general solution to this equation is

$$\psi(x) = Ce^{\kappa x} + De^{-\kappa x}, \tag{4.13}$$

in which C and D are constants to be determined by the boundary conditions.

Outside the infinite rectangular well, where $V = \infty$, the constant κ is infinitely large, which means that both constants C and D must be zero to avoid an infinite-amplitude wavefunction. To understand why that's true, consider what happens for any positive value of x. Since κ is infinitely large, the first term of Eq. 4.13 will also be infinitely large unless $C = 0$, and the exponential factor in the second term will effectively be zero when x is positive and κ is infinitely large. Similarly, for any negative value of x, the second term in Eq. 4.13 will be infinitely large unless $D = 0$, and the first term will effectively be zero. And if both terms of Eq. 4.13 are zero for both positive and negative values of x, the wavefunction $\psi(x)$ must be zero everywhere outside the well.

Since the probability density is equal to the square magnitude of the wavefunction $\psi(x)$, this means that there is zero probability of measuring the position of the particle to be outside the infinite rectangular well. Note that this is *not* true for a finite rectangular well, which you can read about in the next section of this chapter.

Inside the infinite rectangular well, the wavefunction solution to the Schrödinger equation is given by Eq. 4.10, and applying boundary conditions is straightforward. Since the wavefunction $\psi(x)$ must be continuous and must have zero amplitude outside the well, you can set $\psi(x) = 0$ at both the left wall ($x = 0$) and the right wall ($x = a$). At the left wall, $\psi(0) = 0$, so

$$\begin{aligned} \psi(0) = Ae^{ik(0)} + Be^{-ik(0)} &= 0 \\ A + B &= 0 \\ A &= -B. \end{aligned} \tag{5.3}$$

At the right wall, $\psi(a) = 0$, so

$$\begin{aligned} \psi(a) &= Ae^{ika} - Ae^{-ika} = 0 \\ A\left(e^{ika} - e^{-ika}\right) &= 0 \\ \left(e^{ika} - e^{-ika}\right) &= 0, \end{aligned} \tag{5.4}$$

in which the final equation must be true to prevent the necessity of setting $A = 0$, which would result in no wavefunction inside the well.

Using the inverse Euler relation for sine from Chapter 4,

$$\sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i} \tag{4.23}$$

makes Eq. 5.4

$$\left(e^{ika} - e^{-ika}\right) = 2i\sin(ka) = 0. \tag{5.5}$$

This can be true only if ka equals zero or an integer multiple of π. But if $ka = 0$, for any nonzero value of a, then k must be zero. That means that the separation constant E in the Schrödinger equation must also be zero, which means that the wavefunction solution will have zero curvature. Since the boundary conditions at the wall of the infinite rectangular well require that $\psi(0) = \psi(a) = 0$, a wavefunction with no curvature would have zero amplitude everywhere inside (and outside) the well, which would mean the particle would not exist.

So $ka = 0$ is not a good option, and you must choose the alternative approach of making $\sin(ka) = 0$. That means that ka must equal an integer multiple of π, and denoting the multiplicative integer by n makes this

$$ka = n\pi$$

or

$$k_n = \frac{n\pi}{a}, \tag{5.6}$$

in which the subscript n is an indicator that k takes on discrete values.

This is a significant result, because it means that the wavenumbers associated with the energy eigenfunctions within the infinite rectangular well are *quantized*, meaning that they have a discrete set of possible values. In other words, since the boundary conditions require that the wavefunction must have zero amplitude at both edges of the well, the only allowed wavefunctions are those with an integer number of half-wavelengths within the well.

And since the wavenumbers associated with the wavefunctions (the energy eigenfunctions) are quantized, Eq. 4.9 tells you that the allowed energies (the energy eigenvalues) within the well must also be quantized. Those discrete allowed energies can be found by solving Eq. 5.2 for energy:

$$E_n = \frac{k_n^2 \hbar^2}{2m} = \frac{n^2 \pi^2 \hbar^2}{2ma^2}. \tag{5.7}$$

So even before considering the details of the wavefunction $\psi(x)$, the probability density $\psi^* \psi$, or the evolution of $\Psi(x, t)$ over time, a fundamental difference between classical and quantum mechanics has become evident. Just by applying the boundary conditions at the edges of the infinite potential well, you can see that a quantum particle in an infinite rectangular well can take on only certain energies, and the energy of even the lowest-energy state is not zero (this minimum value of energy is called the "zero-point energy").

It's important to realize that the wavenumbers (k_n) are associated with the eigenvalues of the total-energy operator and cannot generally be used to determine the momentum of a particle in a rectangular well using de Broglie's relation ($p = \hbar k$). That's because the eigenfunctions of the total-energy operator are not the same as the eigenfunctions of the momentum operator. As described in Chapter 3, making an energy measurement causes the particle's wavefunction to collapse to an energy eigenfunction, and a subsequent measurement of momentum will cause the particle's wavefunction to collapse to an eigenfunction of the momentum operator. Hence you cannot predict the outcome of the momentum measurement using the energy E_n found in the first measurement, and its associated wavenumber k_n. And as you'll see later in this section, the momentum probability density for a particle in an infinite rectangular well is a continuous function rather than a set of discrete values, although for large values of n the probability density is greatest near the values of $p = \hbar k_n$.

With that caveat in mind, additional insight can be gained by inserting k_n into the TISE solution $\psi(x)$:

$$\psi_n(x) = A\left(e^{ik_n x} - e^{-ik_n x}\right) = A' \sin\left(\frac{n\pi x}{a}\right), \tag{5.8}$$

in which the factor of $2i$ has been absorbed into the leading constant A', and the subscript n represents the quantum number designating the wavenumber k_n and energy level E_n associated with the wavefunction $\psi_n(x)$.

When you're working with quantum wavefunctions, it's generally a good idea to normalize the wavefunction under consideration. That way, you can be certain that the total probability of finding the particle somewhere in space

(in this case, between the boundaries of the infinite rectangular well) is unity. For the wavefunction of Eq. 5.8, normalization looks like this:

$$1 = \int_{-\infty}^{\infty} [\psi_n(x)]^* [\psi_n(x)] dx = \int_0^a \left[A' \sin\left(\frac{n\pi x}{a}\right)\right]^* \left[A' \sin\left(\frac{n\pi x}{a}\right)\right] dx$$
$$= \int_0^a |A'|^2 \sin^2\left(\frac{n\pi x}{a}\right) dx.$$

Since A' is a constant, it comes out of the integral, and the integral can be evaluated using

$$\int \sin^2(cx) dx = \frac{x}{2} - \frac{\sin(2cx)}{4c}.$$

So

$$1 = |A'|^2 \int_0^a \sin^2\left(\frac{n\pi x}{a}\right) dx = |A'|^2 \left[\frac{x}{2} - \frac{\sin\left(\frac{2n\pi x}{a}\right)}{4\frac{n\pi}{a}}\right]\Bigg|_0^a = |A'|^2 \frac{a}{2},$$

which means

$$|A'|^2 = \frac{2}{a}$$

or

$$A' = \sqrt{\frac{2}{a}}.$$

If you're concerned about the negative square root of $|A'|^2$, note that $-A'$ may be written as $A'e^{i\pi}$, and a factor such as $e^{i\theta}$ is called a "global phase factor." That's because it affects only the phase, not the amplitude, of $\psi(x)$, and it applies equally to each of the component wavefunctions that make up $\psi(x)$. Global phase factors cannot have an effect on the probability of any measurement result, since they cancel when the product $\psi^*\psi$ is taken. So no information is lost in taking only the positive square root of $|A'|^2$.

Inserting $\sqrt{\frac{2}{a}}$ for A' into Eq. 5.8 gives the normalized wavefunction $\psi_n(x)$ within the infinite rectangular well:

$$\psi_n(x) = \sqrt{\frac{2}{a}} \sin\left(\frac{n\pi x}{a}\right). \tag{5.9}$$

You can see these wavefunctions $\psi_n(x)$ in Fig. 5.2 for quantum numbers $n = 1, 2, 3, 4$, and 20. Note that the wavefunction $\psi_n(x)$ with the lowest energy level $E_1 = \frac{\pi^2\hbar^2}{2ma^2}$ is often called the "ground state" and has a single half-cycle across the width (a) of the well. This ground-state wavefunction has a node

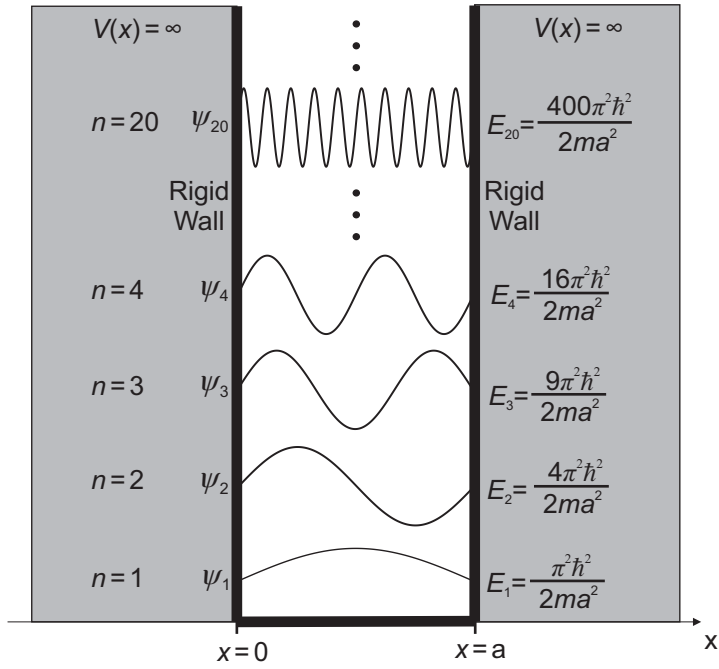


Figure 5.2 Wavefunctions $\psi(x)$ in an infinite rectangular well.

(location of zero amplitude) at each boundary of the well, but no nodes within the well. For higher-energy wavefunctions, often called "excited states," each step up in energy level adds another half-cycle to the wavefunction across the well and another node within the well. So $\psi_2(x)$ has two half-cycles across the well and one node within the well, $\psi_3(x)$ has three half-cycles across the well and two nodes within the well, and so forth.

You should also take a careful look at the symmetry of the wavefunctions with respect to the center of the well. Recall that an even function has the same value at equal distances to the left and to the right of $x = 0$, so $f(x) = f(-x)$. But for an odd function the values of the function on opposite sides of $x = 0$ have opposite signs, so $f(x) = -f(-x)$. As you can see in Fig. 5.3, the wavefunctions $\psi(x)$ alternate between even and odd parity if the center of the well is taken as $x = 0$. Remember that there are many functions that have neither even nor odd parity, in which case $f(x)$ does not equal $f(-x)$ or $-f(-x)$. But one consequence of the form of the Schrödinger equation is that wavefunction solutions will always have either even or odd parity about some point whenever the potential-energy function $V(x)$ is symmetric about

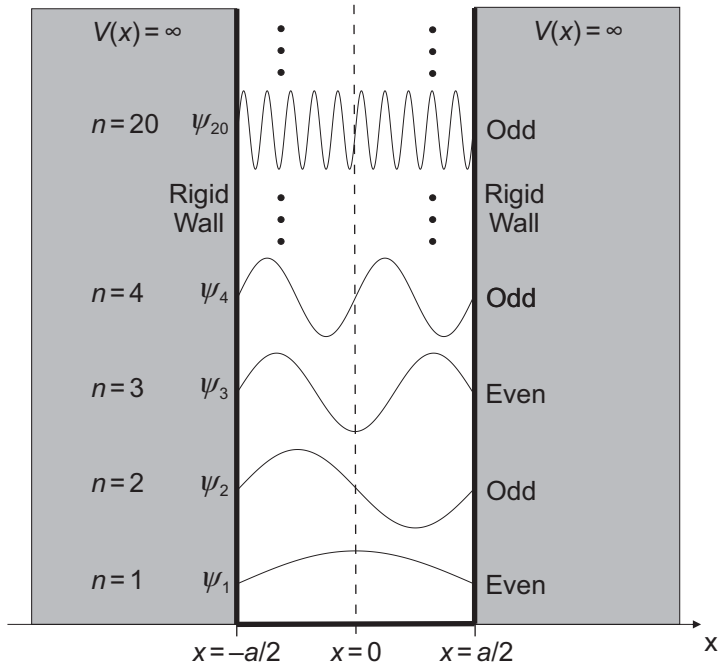


Figure 5.3 Infinite rectangular potential well centered on $x = 0$.

that point, which it is in the case of the infinite rectangular well. For some problems, this definite parity can be helpful in finding solutions, as you'll see in Section 5.2 about finite rectangular wells.

Note also that although the wavefunctions $\psi_n(x)$ are drawn with equal vertical spacing in Figs. 5.2 and 5.3, the energy difference between adjacent wavefunctions increases with increasing n. So the energy-level difference between $\psi_2(x)$ and $\psi_1(x)$ is

$$E_2 - E_1 = \frac{4\pi^2\hbar^2}{2ma^2} - \frac{\pi^2\hbar^2}{2ma^2} = \frac{3\pi^2\hbar^2}{2ma^2},$$

while the energy-level difference between $\psi_3(x)$ and $\psi_2(x)$ is greater:

$$E_3 - E_2 = \frac{9\pi^2\hbar^2}{2ma^2} - \frac{4\pi^2\hbar^2}{2ma^2} = \frac{5\pi^2\hbar^2}{2ma^2}.$$

In general, the spacing between any energy level E_n and the next higher level E_{n+1} is given by

$$E_{n+1} - E_n = (2n+1)\frac{\pi^2\hbar^2}{2ma^2}. \tag{5.10}$$

At this point, it's worthwhile to step back and consider how the Schrödinger equation and the boundary conditions of the infinite rectangular well determine the behavior of the quantum wavefunctions within the well. Remember that the second spatial derivative of $\psi(x)$ on the right side of Eq. 4.8 represents the curvature of the wavefunction, and the separation constant E represents the total energy of the particle. So it's inescapable that higher energy means larger curvature, and for sinusoidally varying functions, larger curvature means more cycles within a given distance (higher value of wavenumber k, so shorter wavelength λ).

Now consider the requirement that the amplitude of the wavefunction must be zero at the edges of the potential well (where $V = \infty$), which means that the distance across the well must correspond to an integer number of half-wavelengths. In light of these conditions, it makes sense that the wavenumber and the energy can take on only those values that cause the curvature of the wavefunction to bring its amplitude back to zero at both edges of the well, as illustrated in Fig. 5.4.

If these infinite-well wavefunctions look familiar to you, it's probably because you've seen the standing waves that are the "normal modes" of vibration of a uniform string rigidly clamped at both ends. For those standing waves, the shape of the wavefunction at the lowest (fundamental) frequency is one half-sinusoid, with one antinode (location of maximum displacement) in the middle, and no nodes except for the two at the locations of the clamped ends of the string. And just as in the case of a quantum particle in an infinite rectangular well, the wavenumbers and allowed energies of the standing waves on a clamped string are quantized, with each higher-frequency mode adding one half-cycle across the length of the string.

The analogy, however, is not perfect, since the Schrödinger equation takes the form of a diffusion equation (with a first-order time derivative and second-order spatial derivative) rather than a classical wave equation (with second-order time and spatial derivatives). For the waves on a uniform string, the angular frequency ω is linearly proportional to the wavenumber k, while in the quantum case $E = \hbar\omega$ is proportional to k^2, as shown in Eq. 5.7. This difference in dispersion relation means that the behavior of quantum wavefunctions over time will differ from the behavior of mechanical standing waves on a uniform string.

You can read about the time evolution of $\Psi(x,t)$ for a particle in an infinite rectangular well later in this section, but first you should consider what the wavefunction $\psi(x)$ tells you about the probable outcome of measurements of observable quantities such as energy (E), position (x), or momentum (p).

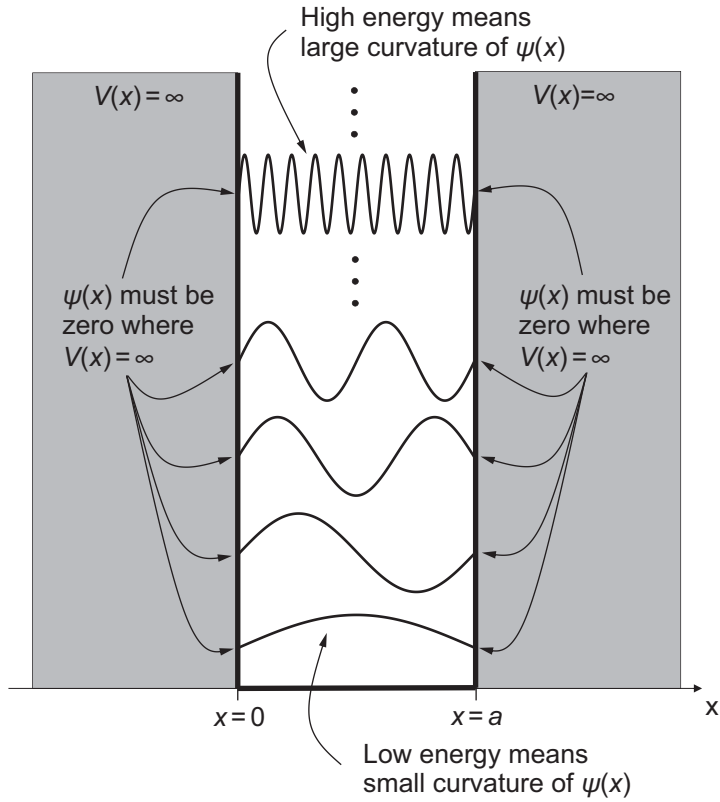


Figure 5.4 Characteristics of wavefunctions $\psi(x)$ in an infinite rectangular well.

As described in Section 3.3, the TISE is an eigenvalue equation for the Hamiltonian (total energy) operator. That means that the wavefunction solutions $\psi_n(x)$ given by Eq. 5.9 are the position-space representation of the energy eigenfunctions, and the energy values given by Eq. 5.7 are the corresponding eigenvalues.

Knowing the eigenfunctions and eigenvalues of the energy operator makes it straightforward to determine the possible outcomes of energy measurements of a quantum particle in an infinite rectangular well. If the particle's state corresponds to one of the eigenfunctions of the Hamiltonian operator (the $\psi_n(x)$ of Eq. 5.9), an energy measurement is certain to yield the eigenvalue of that eigenfunction (the E_n of Eq. 5.7).

And what if the particle's state ψ does not correspond to one of the energy eigenfunctions $\psi_n(x)$? In that case, remember that the eigenfunctions of the

total energy operator form a complete set, so they can be used as basis functions to synthesize any function:

$$\psi = \sum_{n=1}^{\infty} c_n \psi_n(x), \tag{5.11}$$

in which c_n represents the amount of each eigenfunction $\psi_n(x)$ contained in ψ.

This is a version of the Dirac notation Eq. 1.35 from Section 1.6 in which a quantum state represented by ket $|\psi\rangle$ was expanded using eigenfunctions represented by kets $|\psi_n\rangle$:

$$|\psi\rangle = c_1 |\psi_1\rangle + c_2 |\psi_2\rangle + \cdots + c_N |\psi_N\rangle = \sum_{n=1}^{N} c_n |\psi_n\rangle .$$

Recall from Chapter 4 that each c_n may be found by using the inner product to project state $|\psi\rangle$ onto the corresponding eigenfunction $|\psi_n\rangle$,

$$c_n = \langle \psi_n | \psi \rangle . \tag{4.1}$$

Hence for a particle in state ψ in an infinite rectangular well, the "amount" c_n of each eigenfunction $\psi_n(x)$ in that state can be found using the inner product:

$$c_n = \int_0^a [\psi_n(x)]^* [\psi] dx. \tag{5.12}$$

With the values of c_n in hand, you can determine the probability of each measurement outcome (that is, the probability of occurrence of each eigenvalue associated with one of the energy eigenfunctions) by taking the square of the magnitude of the c_n corresponding to that eigenfunction. Over a large ensemble of identically prepared systems, the expectation value of the energy can be found using

$$\langle E \rangle = \sum_n |c_n|^2 E_n. \tag{5.13}$$

If you'd like work through an example of this process, check out the problems at the end of this chapter and the online solutions.

To determine the probable outcomes of position measurements, start by finding the position probability density $P_{den}(x)$ by multiplying the wavefunction $\psi_n(x)$ by its complex conjugate:

$$\begin{aligned} P_{den}(x) &= [\psi(x)]^* [\psi(x)] \\ &= \left[\sqrt{\frac{2}{a}} \sin\left(\frac{n\pi x}{a}\right) \right]^* \left[\sqrt{\frac{2}{a}} \sin\left(\frac{n\pi x}{a}\right) \right] = \frac{2}{a} \sin^2\left(\frac{n\pi x}{a}\right). \end{aligned} \tag{5.14}$$

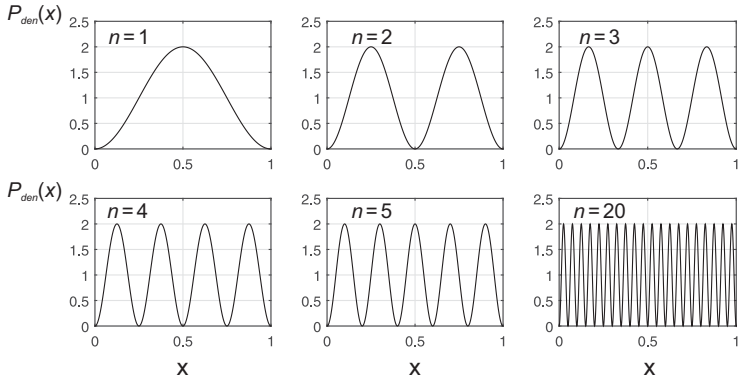


Figure 5.5 Infinite potential well position probability densities.

You can see the position probability densities as a function of x for $n = 1$ through 5 and for $n = 20$ in Fig. 5.5, and they tell an interesting story.

In this figure, each horizontal axis represents distance from the left boundary normalized by the width a of the rectangular well, so the center of the well is at $x = 0.5$ in each plot. Each vertical axis represents probability per unit length, with one length unit defined as the width of the well. As you can see, the probability density $P_{den}(x)$ is a continuous function of x and is not quantized, so a position measurement can yield a value anywhere within the potential well (although for each of the excited states ($n > 1$), there exist one or more positions within the well with zero probability). For a particle in the ground state, the probability density is maximum at the center of the potential well and decreases to zero at the location of the rigid walls. But for the excited states, there are alternating locations of high and low probability density across the well. Thus a position measurement of a particle in the first excited state ($n = 2$) will never result in value of $x = 0.5a$, and a measurement of the position of a particle in the second excited state ($n = 3$) will never result in a value of $x = a/3$ or $x = 2a/3$.

If you integrate the position probability density over the entire potential well, you can be sure of getting a total probability of 1.0, irrespective of the state of the particle. That's because the particle is guaranteed to exist somewhere in the well, so the area under each of the curves in Fig. 5.5 is unity. But if you wish to determine the probability of measuring the particle's position to be within a specified region inside the well, integrate the probability density over that region. For example, to determine the probability

of measuring the particle to be within a region of width Δx centered on position x_0, you can use

$$\int_{x_0-\Delta x/2}^{x_0+\Delta x/2} [\psi(x)]^*[\psi(x)]dx = \int_{x_0-\Delta x/2}^{x_0+\Delta x/2} \frac{2}{a}\sin^2\left(\frac{n\pi x}{a}\right)dx. \tag{5.15}$$

You can see an example of this calculation in the chapter-end problems and online solutions.

To appreciate the significance of the preceding discussion of energy and position observations of a quantum particle in an infinite rectangular well, consider the behavior of a classical object trapped in a zero-force region bounded by infinite, inward-pointing forces. That classical particle may be found to be at rest, with zero total energy, or it may be found to be moving with any constant value of energy (and so at constant speed) in either direction as it travels between the rigid walls. If a position measurement is made, the classical particle is equally likely to be found at any position within the well.

But for a quantum particle bound in an infinite rectangular well, an energy measurement can yield only certain allowed values, specifically the values E_n given in Eq. 5.7, none of which is zero. And the probable results of position measurements depend on the particle's state, but in no case is the probability uniform across the well. Most surprisingly, for particles in an excited state, there are one or more locations with zero probability of occurrence as the output of a position measurement.

So you clearly shouldn't think of a quantum particle in an infinite rectangular well as a very small version of a classical object bouncing back and forth between perfectly reflecting walls. But you can gain additional insight into the particle's behavior by considering the possible outcome of measurements of another observable, and that observable is momentum.

To determine the probable outcomes of momentum measurements for a particle in an infinite rectangular well, you need to know the probability density in momentum space, which you can find using the particle's momentum-space wavefunction $\tilde{\phi}(p)$. As described in Section 4.4, you can find that momentum-basis wavefunction by taking the Fourier transform of $\psi(x)$:

$$\begin{aligned}\tilde{\phi}(p) &= \frac{1}{\sqrt{2\pi\hbar}}\int_{-\infty}^{\infty}\psi(x)e^{-ipx/\hbar}dx \\ &= \frac{1}{\sqrt{2\pi\hbar}}\int_0^a \sqrt{\frac{2}{a}}\sin\left(\frac{n\pi x}{a}\right)e^{-ipx/\hbar}dx \\ &= \frac{1}{\sqrt{\pi a\hbar}}\int_0^a \sin\left(\frac{n\pi x}{a}\right)e^{-ipx/\hbar}dx.\end{aligned}$$

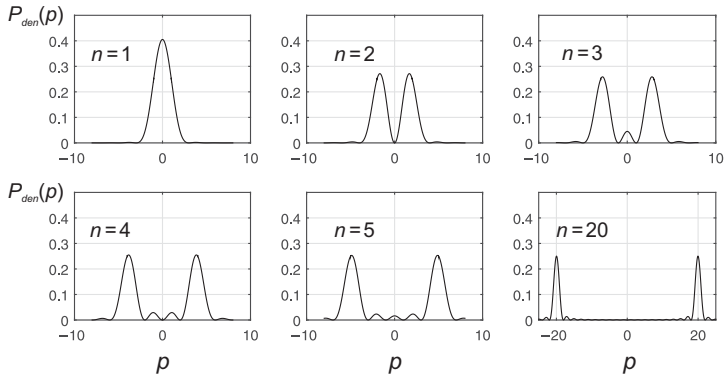


Figure 5.6 Infinite potential well momentum probability densities.

This integral may be evaluated either by using Euler's relation to convert the exponential into sine and cosine terms or by using the inverse Euler relation to convert the sine term into a sum of two exponentials. Either way, the result of the integration is

$$\tilde{\phi}(p) = \frac{\sqrt{\hbar}}{2\sqrt{\pi a}} \left[\frac{2p_n}{p_n^2 - p^2} - \frac{e^{-i(p_n+p)a/\hbar}}{p_n + p} - \frac{e^{i(p_n-p)a/\hbar}}{p_n - p} \right], \tag{5.16}$$

in which $p_n = \hbar k_n$.

The probability density $P_{den}(p)$ can be found by multiplying by $[\tilde{\phi}(p)]^*$, which gives[2]

$$P_{den}(p) = \frac{\hbar}{\pi a} \frac{2p_n^2}{(p_n^2 - p^2)^2} [1 - (-1)^n \cos(pa/\hbar)]. \tag{5.17}$$

In this form, the behavior of the momentum probability density isn't exactly transparent, but you can see plots of $P_{den}(p)$ for the ground state ($n = 1$) and excited states $n = 2, 3, 4, 5$, and 20 in Fig. 5.6. In this figure, each horizontal axis represents normalized momentum (that is, momentum divided by $\hbar\pi/a$), and each vertical axis represents momentum probability density (that is, probability per unit momentum, with one momentum unit defined[3] as $\hbar\pi/a$).

[2] If you need help getting $\tilde{\phi}(p)$ or $P_{den}(p)$, see the chapter-end problems and online solutions.
[3] The reason for this choice of normalization constant will become clear when you consider the most likely values for p for large quantum numbers.

As shown in the plot for $n = 1$, the most likely result of a measurement of the momentum of a quantum particle in the ground state of an infinite rectangular well is $p = 0$, but there is nonzero probability that the measurement will result in a slightly negative or positive value for p. This shouldn't be surprising, since the particle's position is confined to the width a of the potential well, and Heisenberg's Uncertainty principle tells you that $\Delta x \Delta p$ must be equal to or greater than $\hbar/2$. Taking Δx as 18% of the well width a (you can see the reason for this choice in the chapter-end problems and online solutions), and estimating the width of the momentum probability density function as one unit (so $\Delta p = \hbar\pi/a$), the product $\Delta x \Delta p = 0.57\hbar$, so the requirements of the Heisenberg Uncertainty principle $\Delta x \Delta p \geq \hbar/2$ are satisfied.

For excited ($n > 1$) states, the momentum probability density has two peaks, one with positive momentum and another with negative momentum. These correspond to waves traveling in opposite directions, and the larger the quantum number, the closer the probability-density maxima get to $\pm\hbar k_n$, where k_n are the quantized wavenumbers associated with the energy eigenvalues E_n. This is the reason for using $\hbar\pi/a$ as the normalizing factor for momentum; for the lowest-energy state (the ground state), the energy eigenfunction has one half-cycle across the width of the potential well, so the wavelength $\lambda_1 = 2a$, and the wavenumber k_1 associated with that energy has value $k_1 = 2\pi/\lambda_1 = 2\pi/(2a) = \pi/a$. If you were to use deBroglie's relation $p = \hbar k$ to find the momentum associated with that wavenumber, you'd get $\hbar\pi/a$. That's a convenient factor, since the most-likely momentum values cluster around $p = n\hbar\pi/a$ for large values of n, but as you can see in Fig. 5.6, that does not mean that $p_1 = (1)\hbar\pi/a$ is a good estimate of the probable outcome of a momentum measurement for a particle in the ground state of an infinite rectangular well.[4]

As the preceding discussion shows, many aspects of the behavior of a quantum particle in an infinite rectangular well can be understood by solving the TISE and applying the appropriate boundary conditions. But to determine how the particle's wavefunction evolves over time, it's necessary to consider the solutions to the time-dependent Schrödinger equation, and

[4] In fact, the probability density function for the ground state does consist of two component functions, one of which has a peak at $+\hbar k_1 = +\hbar\pi/a$, while the other has a peak at $-\hbar k_1 = -\hbar\pi/a$. But the width of those two functions is sufficiently great to cause them to overlap, and they combine to produce the peak at $p = 0$ in the ground-state probability-density function.

that means including the time portion $T(t)$ from the separation of variables ($T(t) = e^{-iE_n t/\hbar}$):

$$\Psi_n(x,t) = \psi_n(x)T(t) = \sqrt{\frac{2}{a}} \sin\left(\frac{n\pi x}{a}\right) e^{-iE_n t/\hbar}. \tag{5.18}$$

You may recall from Section 3.3 that separable solutions to the time-dependent Schrödinger equation are called "stationary states" because quantities such as expectation values and probability density functions associated with such states do not vary over time. That certainly pertains to the eigenfunctions $\Psi_n(x,t)$ given by Eq 5.18 for the Hamiltonian operator of the infinite rectangular well, since for any given value of n the exponential term $e^{-iE_n t/\hbar}$ will cancel out when $\Psi_n(x,t)$ is multiplied by its complex conjugate. That means that for a particle in any of the energy eigenstates of the infinite rectangular well, the position- and momentum-based probability densities shown in Figs. 5.5 and 5.6 will not change as time passes.

The situation is quite different if a particle in an infinite rectangular well is in a state that is not an energy eigenstate. Consider, for example, a particle in a state that has a wavefunction that is the linear superposition of the first and second energy eigenfunctions:

$$\begin{aligned}\Psi(x,t) &= A\Psi_1(x,t) + B\Psi_2(x,t) \\ &= A\sqrt{\frac{2}{a}} \sin\left(\frac{\pi x}{a}\right) e^{-iE_1 t/\hbar} + B\sqrt{\frac{2}{a}} \sin\left(\frac{2\pi x}{a}\right) e^{-iE_2 t/\hbar},\end{aligned} \tag{5.19}$$

in which the constants A and B determine the relative amounts of eigenfunctions $\Psi_1(x,t)$ and $\Psi_2(x,t)$. Note that the leading factor $\sqrt{\frac{2}{a}}$ in Eq. 5.18, which was determined by normalizing individual eigenfunctions $\Psi_n(x,t)$, is not the correct normalizing factor when two or more eigenfunctions are combined. So in addition to setting the relative amounts of the constituent eigenfunctions, the factors A and B will also provide the proper normalization for the composite function $\Psi(x,t)$.

For example, to synthesize a total wavefunction consisting of equal parts of the infinite rectangular well eigenfunctions Ψ_1 and Ψ_2, the factors A and B must be equal. The normalization process is:

$$\begin{aligned}1 &= \int_{-\infty}^{+\infty} \Psi^*\Psi dx = \int_{-\infty}^{+\infty} [A\Psi_1 + A\Psi_2]^*[A\Psi_1 + A\Psi_2]dx \\ &= |A|^2 \int_{-\infty}^{+\infty} [\Psi_1^*\Psi_1 + \Psi_1^*\Psi_2 + \Psi_2^*\Psi_1 + \Psi_2^*\Psi_2]dx.\end{aligned}$$

Plugging in Ψ_1 and Ψ_2 from Eq. 5.19 with $A = B$ makes this

$$
\begin{aligned}
1 = |A|^2 \Bigg\{ & \int_0^a \left[\sqrt{\frac{2}{a}} \sin\left(\frac{\pi x}{a}\right) e^{-iE_1 t/\hbar} \right]^* \left[\sqrt{\frac{2}{a}} \sin\left(\frac{\pi x}{a}\right) e^{-iE_1 t/\hbar} \right] dx \\
& + \int_0^a \left[\sqrt{\frac{2}{a}} \sin\left(\frac{\pi x}{a}\right) e^{-iE_1 t/\hbar} \right]^* \left[\sqrt{\frac{2}{a}} \sin\left(\frac{2\pi x}{a}\right) e^{-iE_2 t/\hbar} \right] dx \\
& + \int_0^a \left[\sqrt{\frac{2}{a}} \sin\left(\frac{2\pi x}{a}\right) e^{-iE_2 t/\hbar} \right]^* \left[\sqrt{\frac{2}{a}} \sin\left(\frac{\pi x}{a}\right) e^{-iE_1 t/\hbar} \right] dx \\
& + \int_0^a \left[\sqrt{\frac{2}{a}} \sin\left(\frac{2\pi x}{a}\right) e^{-iE_2 t/\hbar} \right]^* \left[\sqrt{\frac{2}{a}} \sin\left(\frac{2\pi x}{a}\right) e^{-iE_2 t/\hbar} \right] dx \Bigg\}.
\end{aligned}
$$

Note that the limits of integration are now 0 to a, since the wavefunctions have zero amplitude outside that region. Carrying out the multiplications gives

$$
\begin{aligned}
1 = |A|^2 \left(\frac{2}{a}\right) \Bigg\{ & \int_0^a \left[\sin^2\left(\frac{\pi x}{a}\right) \right] dx \\
& + \int_0^a \left[\sin\left(\frac{\pi x}{a}\right) \right] \left[\sin\left(\frac{2\pi x}{a}\right) \right] e^{-i(E_2 - E_1)t/\hbar} dx \\
& + \int_0^a \left[\sin\left(\frac{2\pi x}{a}\right) \right] \left[\sin\left(\frac{\pi x}{a}\right) \right] e^{+i(E_2 - E_1)t/\hbar} dx \\
& + \int_0^a \left[\sin^2\left(\frac{2\pi x}{a}\right) \right] dx \Bigg\}.
\end{aligned}
$$

Just as in the normalization process shown earlier for the individual eigenfunctions, the first and last of these integrals each give a value of $\frac{a}{2}$, so those two terms add up to a. As for the cross terms (the second and third integrals), the orthogonality of the energy eigenfunctions means that each of these integrals gives zero, so $1 = |A|^2 \left(\frac{2}{a}\right)(a)$, which means $A = \frac{1}{\sqrt{2}}$ in the case of equal amounts of $\Psi_1(x,t)$ and $\Psi_2(x,t)$.

A similar analysis for any two factors A and B indicates that the composite function will be properly normalized as long as $|A|^2 + |B|^2 = 1$. So for a mixture in which the amount A of $\Psi_1(x,t)$ is 0.96, the amount B of $\Psi_2(x,t)$ must equal 0.28 (since $0.96^2 + 0.28^2 = 1$).

One reason for considering such a lopsided combination of $\Psi_1(x,t)$ and $\Psi_2(x,t)$ is to demonstrate that adding even a small amount of a different eigenfunction into the mix can have a significant effect on the behavior of the composite wavefunction. You can see that effect in Fig. 5.7, which

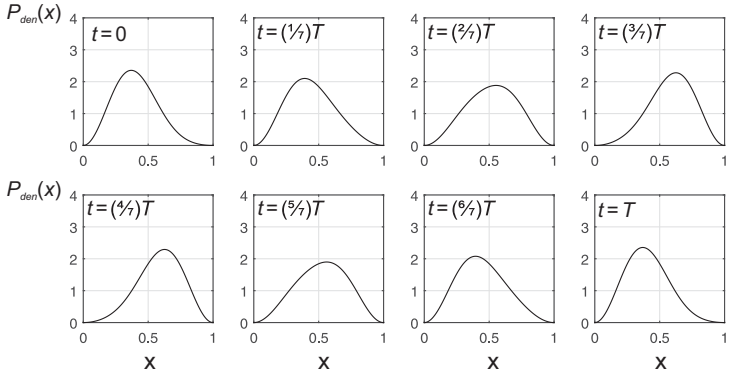


Figure 5.7 Time evolution of position probability density in infinite rectangular well for a mixture with $(0.96)\Psi_1(x,t)$ and $(0.28)\Psi_2(x,t)$.

shows the time evolution of the position probability density of the composite wavefunction

$$\Psi(x,t) = (0.96)\sqrt{\frac{2}{a}}\sin\left(\frac{\pi x}{a}\right)e^{-iE_1t/\hbar} + (0.28)\sqrt{\frac{2}{a}}\sin\left(\frac{2\pi x}{a}\right)e^{-iE_2t/\hbar}. \tag{5.20}$$

As you can see, the position probability is no longer stationary – the mixture of energy eigenfunctions causes the position of maximum probability density to oscillate within the infinite rectangular well. For a mixture with a heavy dose of $\Psi_1(x,t)$, the shape of the probability density function resembles the single-peaked shape of the probability density of $\Psi_1(x,t)$, but the presence of a small amount of $\Psi_2(x,t)$ with its double-peaked probability density has the effect of sliding the probability density of the composite function back and forth as the two constituent eigenfunctions cycle in and out of phase with one another.

And why does that happen? Because the energies of $\Psi_1(x,t)$ and $\Psi_2(x,t)$ are different, and energy is related to angular frequency by the Planck–Einstein relation $E = hf = \hbar\omega$ (Eq. 3.1), as discussed in Chapter 3. So different energies mean different frequencies, and different frequencies mean that as time passes, the relative phase between $\Psi_1(x,t)$ and $\Psi_2(x,t)$ changes. That changing phase causes various parts of these two wavefunctions to add or subtract, changing the shape of the composite wavefunction and its probability density function.

The mathematics of that phase variation is not difficult to comprehend. The terms of the product $[\Psi(x,t)]^*[\Psi(x,t)]$ appear as the integrands of the

normalization integrals shown earlier for the case in which the amounts of the two wavefunctions are equal (that is, $A = B$). In the general case, A and B may have different values, and the terms of $[\Psi(x,t)]^*[\Psi(x,t)]$ are given by

$$\begin{aligned} P_{den}(x,t) = {} & |A|^2 \left(\frac{2}{a}\right)\left[\sin^2\left(\frac{\pi x}{a}\right)\right] + |B|^2 \left(\frac{2}{a}\right)\left[\sin^2\left(\frac{2\pi x}{a}\right)\right] \\ & + |A||B| \left(\frac{2}{a}\right)\left[\sin\left(\frac{\pi x}{a}\right)\right]\left[\sin\left(\frac{2\pi x}{a}\right)\right] e^{-i(E_2-E_1)t/\hbar} \\ & + |A||B| \left(\frac{2}{a}\right)\left[\sin\left(\frac{2\pi x}{a}\right)\right]\left[\sin\left(\frac{\pi x}{a}\right)\right] e^{+i(E_2-E_1)t/\hbar}. \end{aligned}$$

The first term, involving only $\Psi_1(x,t)$, and the second term, involving only $\Psi_2(x,t)$, have no time dependence because the exponential term $e^{-iE_nt/\hbar}$ has canceled out in each case. But the different energies of $\Psi_1(x,t)$ and $\Psi_2(x,t)$ mean that the cross terms $[\Psi_1]^*[\Psi_2]$ and $[\Psi_2]^*[\Psi_1]$ retain their time dependence.

You can see the effect of that time dependence by writing the combination of those two cross terms as

$$\begin{aligned} & |A||B| \left(\frac{2}{a}\right)\left[\sin\left(\frac{\pi x}{a}\right)\right]\left[\sin\left(\frac{2\pi x}{a}\right)\right] e^{-i(E_2-E_1)t/\hbar} \\ & \quad + |A||B| \left(\frac{2}{a}\right)\left[\sin\left(\frac{2\pi x}{a}\right)\right]\left[\sin\left(\frac{\pi x}{a}\right)\right] e^{+i(E_2-E_1)t/\hbar} \\ & = |A||B| \left(\frac{2}{a}\right)\left[\sin\left(\frac{\pi x}{a}\right)\right]\left[\sin\left(\frac{2\pi x}{a}\right)\right] [e^{-i(E_2-E_1)t/\hbar} + e^{+i(E_2-E_1)t/\hbar}] \\ & = 2|A||B| \left(\frac{2}{a}\right)\left[\sin\left(\frac{\pi x}{a}\right)\right]\left[\sin\left(\frac{2\pi x}{a}\right)\right] \cos\left[\frac{(E_2-E_1)t}{\hbar}\right]. \end{aligned}$$

The time variation in $P_{den}(x,t)$ is caused by the cosine term, and that term depends on the difference between the energy levels of the two energy eigenfunctions that make up the composite wavefunction $\Psi(x,t)$. The larger the energy difference, the faster the oscillation of that cosine term, as you can see by writing the angular frequency of the composite wavefunction as

$$\omega_{21} = \omega_2 - \omega_1 = \frac{E_2 - E_1}{\hbar} \tag{5.21}$$

or, using the energy levels of the infinite rectangular well (Eq. 5.7),

$$\omega_{21} = \frac{E_2 - E_1}{\hbar} = \frac{2^2\pi^2\hbar^2}{2ma^2\hbar} - \frac{1^2\pi^2\hbar^2}{2ma^2\hbar} = \frac{3\pi^2\hbar}{2ma^2}. \tag{5.22}$$

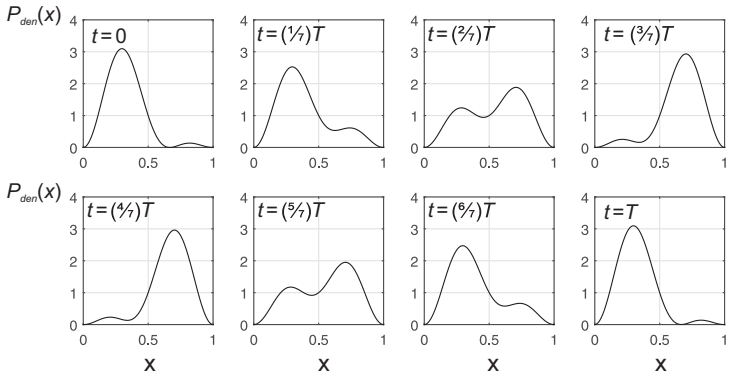


Figure 5.8 Time evolution of position probability density in infinite rectangular well for an equal mixture of $\Psi_1(x,t)$ and $\Psi_2(x,t)$.

As you might expect, adding in a larger amount of $\Psi_2(x,t)$ has the effect of modifying the shape of the composite probability density function more significantly, as shown in Fig. 5.8. In this case, the amount of $\Psi_2(x,t)$ is equal to the amount of $\Psi_1(x,t)$.

Note that the larger proportion of $\Psi_2(x,t)$ causes the peak of the position probability density at time $t = 0$ to occur farther to the left (toward the position of large positive amplitude of $\Psi_2(x,t)$ shown in Fig. 5.2). As time passes, the higher angular frequency of $\Psi_2(x,t)$ causes its phase to change more quickly than the phase of $\Psi_1(x,t)$, shifting the peak of the probability density to the right. After one half-cycle of the composite wavefunction (for which the period T is $\frac{2\pi}{\omega_{21}}$), the probability-density peak has moved into the right half of the rectangular well. After one complete cycle of the composite wavefunction, the peak of the probability density is once again in the left portion of the well.

This analysis shows that for states of the infinite rectangular well that are made up of the weighted combination of eigenstates, the probability density function varies over time. The amount of that variation depends on the relative proportions of the component states, and the rate of the variation is determined by the energies of those states.

Many of the concepts and techniques described in this section are applicable to potential wells with somewhat more realistic potential energy configurations. You can read about one of those configurations, the finite rectangular well, in the next section.

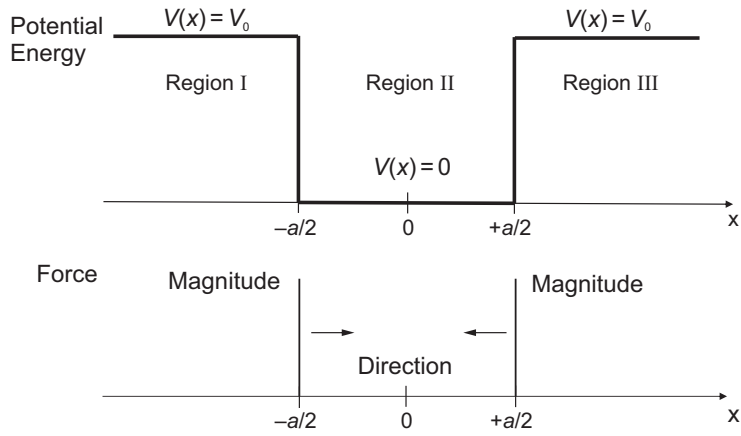


Figure 5.9 Finite potential well energy and force as a function of location.

5.2 Finite Rectangular Potential Well

Like the infinite rectangular potential well, the finite rectangular well is an example of a configuration with piecewise constant potential energy, but in this case the potential energy has a constant finite (rather than infinite) value outside the well. An example of a finite rectangular well is shown in Fig. 5.9, and as you can see, the bottom of the well can be taken as the reference level of zero potential energy, while the potential energy $V(x)$ outside the well has constant value V_0.[5]

You should also note that the width of this finite potential well is taken as a, but the center of the well is shown at location $x = 0$, which puts the left edge of the well at position $x = -a/2$ and the right edge at position $x = a/2$. The location that you choose to call $x = 0$ has no impact on the physics of the finite potential well or the shape of the wavefunctions, but taking $x = 0$ at the center does make the wavefunction parity considerations more apparent, as you'll see presently.

The solutions to the Schrödinger equation within and outside the finite rectangular well have several similarities but a few important differences from those of the infinite rectangular well discussed in the previous section.

[5] Some quantum texts take the reference level of zero potential energy as outside the well, in which case the potential energy at the bottom of the well is $-V_0$. As in classical physics, only the *change* is potential energy has any physical significance, so you're free to choose whichever location is most convenient as the reference level.

Similarities include the oscillatory nature of the wavefunction $\psi(x)$ within the well and the requirement for the value of the wavefunction to be continuous across the walls of the well (that is, at $x = -a/2$ and $x = a/2$). But since the potential energy outside the finite potential well is not infinite, the wavefunction is not required to have zero amplitude outside the well. That means that it's also necessary for the slope of the wavefunction $\frac{\partial \psi(x)}{\partial x}$ to be continuous across the walls. These boundary conditions lead to a somewhat more complicated equation from which the allowed energy levels and wavefunctions may be extracted.

Another important difference between the finite and the infinite potential well is this: for a finite potential well, particles may be bound or free, depending on their energy and the characteristics of the well. Specifically, for the potential energy defined as in Fig. 5.9, the particle will be bound if $E < V_0$ and free if $E > V_0$. In this section, the energy will be taken as $0 < E < V_0$, so the wavefunctions and energy levels will be those of bound particles.

The good news is that if you've worked through Chapter 4, you've already seen the most important features of the finite potential well. That is, the wavefunction solutions are oscillatory inside the well, but they do not go to zero at the edges of the well. Instead, they decay exponentially in that region, often called the "evanescent" region.

And just as in the case of an infinite rectangular well, the wavenumbers and energies of particles bound in a finite rectangular well are quantized (that is, they take on only certain discrete "allowed" values). But for a finite potential well, the number of allowed energy levels is not infinite, depending instead on the width and the "depth" of the well (that is, the difference in potential energy inside and outside of the well).

In this section, you'll find an explanation of why the energy levels are discrete in the finite potential well along with an elucidation of the meaning of the variables used in many quantum texts in the transcendental equation that arises from applying the boundary conditions of the finite rectangular well.

If you've read Section 4.3, you've already seen the basics of wavefunction behavior in a region of piecewise constant potential, in which the total energy E of the quantum particle may be greater than or less than the potential energy V in the region. The curvature analysis presented in that section indicates that wavefunctions in classically allowed regions ($E > V$) exhibit oscillatory behavior, and wavefunctions in classically forbidden regions ($E < V$) exhibit exponentially decaying behavior. Applying these concepts to a quantum particle in a finite rectangular well with potential energy $V = 0$

inside the well and $V = V_0$ outside the well tells you that the wavefunction of a particle with energy $E > 0$ within the well will oscillate sinusoidally.

To see how that result comes about mathematically, write the time-independent Schrödinger equation (Eq. 4.7) inside the well as

$$\frac{d^2[\psi(x)]}{dx^2} = -\frac{2m}{\hbar^2}E\psi(x) = -k^2\psi(x), \tag{5.23}$$

in which the constant k is defined as

$$k \equiv \sqrt{\frac{2m}{\hbar^2}E}, \tag{5.24}$$

exactly as in the case of the infinite rectangular well.

The solution to Eq. 5.23 may be written using either exponentials (as was done in Sections 4.3 and 5.1), or sinusoidal functions. As mentioned in Section 5.1, the wavefunction solutions to the Schrödinger equation will have definite (even or odd) parity whenever the potential-energy function $V(x)$ is symmetric about some point. For the finite rectangular well, this definite parity means that sinusoidal functions are somewhat easier to work with. So the general solution to the Schrödinger equation within the finite rectangular well may be written as

$$\psi(x) = A\cos(kx) + B\sin(kx), \tag{5.25}$$

in which constants A and B are determined by the boundary conditions.

As discussed in Section 4.3, the constant k represents the wavenumber in this region, which determines the wavelength of the quantum wavefunction $\psi(x)$ through the relation $k = 2\pi/\lambda$. Using the logic presented in Chapter 4 relating curvature to energy and wavenumber, Eq. 5.24 tells you that the larger the particle's total energy E, the faster the particle's wavefunction will oscillate with x in a finite rectangular well.

In the regions to the left and right of the potential well the potential energy $V(x) = V_0$ exceeds the total energy E, so the quantity $E - V_0$ is negative, and these are classically forbidden regions. In those regions, the TISE (Eq. 4.7) can be written as

$$\frac{d^2[\psi(x)]}{dx^2} = -\frac{2m}{\hbar^2}(E - V_0)\psi(x) = +\kappa^2\psi(x), \tag{5.26}$$

in which the constant κ is defined as

$$\kappa \equiv \sqrt{\frac{2m}{\hbar^2}(V_0 - E)}. \tag{5.27}$$

Section 4.3 also explains that the constant κ is a "decay constant" that determines the rate at which the wavefunction tends toward zero in a classically forbidden region. And since Eq. 5.27 states that κ is directly proportional to the square root of $V_0 - E$, you know that the greater the amount by which the potential energy V_0 exceeds the total energy E, the larger the decay constant κ, and the faster the wavefunction decays over x (if $V_0 = \infty$ as in the case of the infinite rectangular well, the decay constant is infinitely large, and the wavefunction's amplitude decays to zero at the boundaries of the well).

The general solution to Eq. 5.26 is

$$\psi(x) = Ce^{\kappa x} + De^{-\kappa x}, \tag{5.28}$$

with constants C and D determined by the boundary conditions.

Even before applying the boundary conditions, you can determine something about the constants C and D in the regions outside the finite rectangular well. Calling those constants C_{left} and D_{left} in Region I to the left of the well ($x < -a/2$), the second term of Eq. 5.28 ($D_{left}e^{-\kappa x}$) will become infinitely large unless the coefficient D_{left} is zero in this region. Likewise, calling the constants C_{right} and D_{right} in Region III to the right of the well ($x > a/2$), the first term of Eq. 5.28 ($C_{right}e^{\kappa x}$) will become infinitely large unless C_{right} is zero in this region.

So $\psi(x) = C_{left}e^{\kappa x}$ in Region I where x is negative, and $\psi(x) = D_{right}e^{-\kappa x}$ in Region III where x is positive. And since the symmetry of the potential $V(x)$ about $x = 0$ means that the wavefunction $\psi(x)$ must have either even or odd parity across all values of x (not just within the potential well), you also know that C_{left} must equal D_{right} for even solutions and that C_{left} must equal $-D_{right}$ for odd solutions. So for even solutions you can write $C_{left} = D_{right} = C$, and for odd solutions $C_{left} = C$ and $D_{right} = -C$.

These conclusions about the wavefunction $\psi(x)$ are summarized in the following table, which also shows the first spatial derivative $\frac{\partial\psi(x)}{\partial x}$ of the wavefunction in each of the three regions.

Region:	I	II	III
Behavior:	Evanescent	Oscillatory	Evanescent
$\psi(x)$:	$Ce^{\kappa x}$	$A\cos(kx)$ or $B\sin(kx)$	$Ce^{-\kappa x}$ or $-Ce^{-\kappa x}$
$\frac{\partial\psi(x)}{\partial x}$:	$\kappa Ce^{\kappa x}$	$-kA\sin(kx)$ or $kB\cos(kx)$	$-\kappa Ce^{-\kappa x}$ or $\kappa Ce^{-\kappa x}$

With the wavefunction $\psi(x)$ in hand both inside and outside the well, you're in position to apply the boundary conditions at both the left edge ($x = -a/2$) and the right edge ($x = a/2$) of the finite rectangular well. As in the case of the infinite rectangular well, application of the boundary conditions leads directly to quantization of energy E and wavenumber k for quantum particles within the well.

Considering the even solutions first, matching the amplitude of $\psi(x)$ across the wall at the left edge of the well gives

$$Ce^{\kappa\left(-\frac{a}{2}\right)} = A\cos\left[k\left(-\frac{a}{2}\right)\right] \tag{5.29}$$

and matching the slope (the first spatial derivative) at the left wall gives

$$\kappa Ce^{\kappa\left(-\frac{a}{2}\right)} = -kA\sin\left[k\left(-\frac{a}{2}\right)\right]. \tag{5.30}$$

If you now divide Eq. 5.30 by Eq. 5.29, (forming a quantity called the logarithmic derivative, which is $\frac{1}{\psi}\frac{\partial\psi}{\partial x}$), the result is

$$\frac{\kappa Ce^{\kappa\left(-\frac{a}{2}\right)}}{Ce^{\kappa\left(-\frac{a}{2}\right)}} = \frac{-kA\sin\left[k\left(-\frac{a}{2}\right)\right]}{A\cos\left[k\left(-\frac{a}{2}\right)\right]} \tag{5.31}$$

or

$$\kappa = -k\tan\left(-\frac{ka}{2}\right) = k\tan\left(\frac{ka}{2}\right). \tag{5.32}$$

Dividing both sides by the wavenumber k makes this

$$\frac{\kappa}{k} = \tan\left(\frac{ka}{2}\right). \tag{5.33}$$

This is the mathematical expression of the requirement that the amplitude and slope of the wavefunction $\psi(x)$ must be continuous across the boundary between regions.

To understand why this equation leads to quantized wavenumber and energy levels, recall that the TISE tells you that the constant κ determines the curvature (and thus the decay rate) of $\psi(x)$ in the evanescent regions (I and III), and that the decay constant κ is proportional to the square root of $V_0 - E$. Also note that $V_0 - E$ gives the difference between the particle's energy level and the top of the potential well. So on the evanescent side of the potential well boundaries at $x = -a/2$ and $x = a/2$, the value of $\psi(x)$ and its slope are determined by the "depth" of the particle's energy level in the well.

Now think about the wavenumber k in the oscillatory region. You know from the Schrödinger equation that k determines the curvature (and thus the

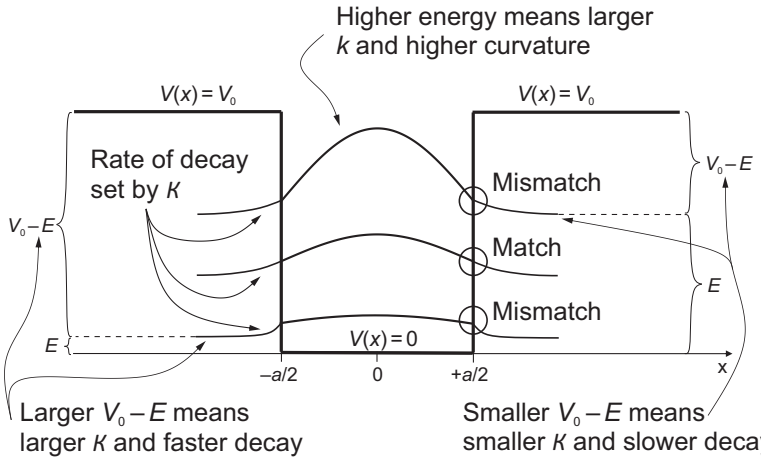


Figure 5.10 Matching slopes in a finite potential well.

spatial rate of oscillation) of $\psi(x)$ in the oscillating region (II), and you also know that k is proportional to the square root of the energy E. But since $V(x) = 0$ at the bottom of the potential well, E is just the difference between the particle's energy level and the bottom of the well. So the value of $\psi(x)$ and its slope on the inside of the potential well boundaries are determined by the "height" of the particle's energy level in the well.

And here's the payoff of this logic: only certain values of energy (that is, certain ratios of the depth to the height of the particle's energy level) will cause both $\psi(x)$ and its first derivative $\frac{\partial \psi(x)}{\partial x}$ to be continuous across the boundaries of the finite potential well. For even solutions, those ratios are given by Eq. 5.33. Sketches of wavefunctions with matching or mismatching slopes at the edges of the well are shown in Fig. 5.10.

Unfortunately, Eq. 5.33 is a **transcendental equation**,[6] which cannot be solved analytically. But with a bit of thought (or perhaps meditation), you can imagine solving this equation either numerically or graphically. The numerical approach is essentially trial and error, hopefully using a clever algorithm to help you guess efficiently. Most quantum texts use some form of graphical approach to solving Eq. 5.33 for the energy levels of the finite rectangular well, so you should make sure you understand how that process works.

[6] A transcendental equation is an equation involving a transcendental function such as a trigonometric or exponential function.

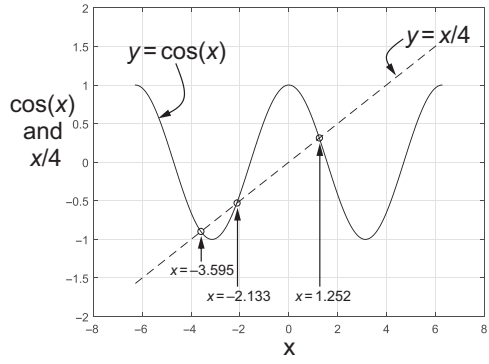


Figure 5.11 Graphical solution of transcendental equation $\frac{x}{4} = \cos(x)$.

To do that, it may help to start by considering a simple transcendental equation, such as

$$\frac{x}{4} = \cos(x). \tag{5.34}$$

The solutions to this equation can be read off the graph shown in Fig. 5.11. As you can see, the trick is to plot both sides of the equation you're trying to solve on the same graph. So in this case, the function $y(x) = \frac{x}{4}$ (from the left side of Eq. 5.34) is plotted on the same set of axes as the function $y(x) = \cos(x)$ (from the right side of Eq. 5.34). This graph makes the solutions to the equation clear: just look for the values of x at which the lines cross, because at those locations $\frac{x}{4}$ must equal $\cos(x)$. In this example, those values are near $x = -3.595$, $x = -2.133$, and $x = +1.252$, and you can verify that these values satisfy the equation by plugging them in to Eq. 5.34 (don't forget to use radians for x, as you must always do when dealing with an angle that appears outside of any trigonometric function, such as the term $x/4$ in Eq. 5.34).

Things are a bit more complex for Eq. 5.33, but the process is the same: plot both sides of the equation on the same graph and look for points of intersection. In many quantum texts, some of the variables are combined and renamed in order to simplify the appearance of the terms in the transcendental equation, but it's been my experience that this can cause students to lose sight of the physics underlying the equation. So before showing you the most common substitution of variables and explaining exactly what the combined variables mean, you may find it instructive to take a look at the graphical solutions for several finite potential wells with specified width and depth.

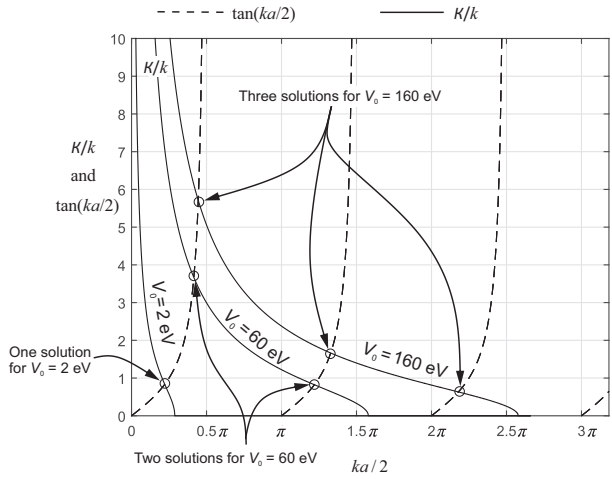


Figure 5.12 Finite potential well graphical solution (even case) for three values of V_0.

In Fig. 5.12, you can see the graphical solution process at work for three finite rectangular wells, all with width $a = 2.5 \times 10^{-10}$ m and with potential energy V_0 of 2, 60, and 160 eV. Plotting the two sides of Eq. 5.33 on the same graph for three different potential wells can make the plot a bit daunting at first glance, but you can understand it by considering the different elements one at a time.

In this graph, the three solid curves represent the ratio κ/k for the three values of V_0. If you're not sure why different values of V_0 give different curves, remember that Eq. 5.27 tells you that κ depends on V_0, so it makes sense that for a given value of k, the value of κ/k is larger for a well with larger potential energy V_0. But each of the three wells has a single value of V_0, so what's changing along each curve? The answer is in the denominator of κ/k, because the horizontal axis of this graph represents a range of values of $ka/2$ (that is, the wavenumber k times the half-width $a/2$ of the wells). As $ka/2$ increases from just above zero to approximately 3π, the ratio κ/k decreases because the denominator is getting bigger.[7]

And where does the total energy E appear on this plot? Remember that you're using this graph to find the allowed energies for each these three well-depths (that is, the energies at which the amplitude and the slope of $\psi(x)$ at the outside edges of the well match the amplitude and the slope at the inside

[7] The graph can't start exactly at $ka/2 = 0$ because that would make the ratio κ/k infinitely large.

edges). To find those allowed values of energy E, you'd like the κ/k curve for each well to run through a range of energy values so you can find the locations (if any) at which the curve intersects $\tan(ka/2)$, shown as dashed curves in Fig. 5.12. At those locations, you can be sure that Eq. 5.33 is satisfied.

This explains why the horizontal axis represents $ka/2$: the wavenumber k is proportional to the square root of the energy E by Eq. 5.24, so a range of $ka/2$ is equivalent to a range of energy values. Thus the three solid curves represent the ratio κ/k over a range of energies, and you can determine those energies from the range of $ka/2$, as you'll later see.

Before determining the energies represented in this graph, take a look at the curve representing κ/k for $V_0 = 160$ eV. That curve intersects the curves representing $\tan(ka/2)$ in three places. So for this finite potential well of width $a = 2.5 \times 10^{-10}$ m and depth of 160 eV, there are three discrete values of the wavenumber k for which the even wavefunctions $\psi(x)$ have amplitudes and slopes that are continuous across the edges of the well (that is, they satisfy the boundary conditions). And three discrete values of wavenumber k mean three discrete values of energy E, in accordance with Eq. 5.24.

Looking at the other two solid curves in Fig. 5.12, you should also note that the 2-eV finite well has a single allowed energy level, while the 60-eV well has two allowed energies. So deeper wells may support more allowed energy levels, but notice the word "may" in this sentence. As you can see in the figure, additional solutions to the transcendental equation occur when the curve representing κ/k intersects additional cycles of the $\tan(ka/2)$ curve. Any increase in well depth shifts the κ/k curve upward, but if that shift isn't large enough to produce another intersection with the next $\tan(ka/2)$ curve (or $-\cot(ka/2)$ odd-solution curve, described later in this section), the number of allowed energy levels will not change. So, in general, deeper wells support more allowed energies (remember that an infinite potential well has an infinite number of allowed energies), but the only way to know the how many energy levels a given well supports is to solve the transcendental equations for both the even-parity and odd-parity solutions to the Schrödinger equation.

To determine the three allowed energies for the 160-eV finite well using Eq. 5.24, in addition to the width a of the well, you also need to know the mass m of the particle under consideration. For this graph, the particle was taken to have the mass of an electron ($m = 9.11 \times 10^{-31}$ kg). Solving Eq. 5.24 for E gives

$$\sqrt{\frac{2m}{\hbar^2}E} = k,$$

or

$$E = \frac{\hbar^2 k^2}{2m}. \tag{5.35}$$

Since the values on the horizontal axis represent $ka/2$ rather than k, it's useful to write this equation in terms of $ka/2$:

$$E = \frac{\hbar^2 \left(\frac{ka}{2}\right)^2 \left(\frac{2}{a}\right)^2}{2m} = 2\frac{\hbar^2 \left(\frac{ka}{2}\right)^2}{ma^2}. \tag{5.36}$$

So for the range of $ka/2$ of approximately 0 to 3π shown in Fig. 5.12, the energy range on the plot extends from $E = 0$ to

$$\begin{aligned} E &= 2\frac{\hbar^2(3\pi)^2}{ma^2} = 2\frac{(1.0546 \times 10^{-34}\ \text{Js})^2(3\pi)^2}{(9.11 \times 10^{-31}\ \text{kg})(2.5 \times 10^{-10}\ \text{m})^2} \\ &= 3.47 \times 10^{-17}\ \text{J} = 216.6\ \text{eV}. \end{aligned}$$

Knowing how to convert values of $ka/2$ to values of energy allows you to perform the final step in determining the allowed energy levels of the finite potential well. That step is to read the $ka/2$ value of each intersection of the κ/k and $\tan(ka/2)$ curves, which you can do by dropping a perpendicular line to the horizontal $(ka/2)$ axis, as shown in Fig. 5.13 for the $V_0 = 160$-eV curve. In this case, the intersections occur (meaning the equation $\kappa/k = \tan(ka/2)$ is satisfied) at $ka/2$ values of 0.445π, 1.33π, and 2.18π. Plugging those values into Eq. 5.36 gives the allowed energy values of 4.76 eV, 42.4 eV, and 114.3 eV.

It's reassuring that none of these values exceeds the depth of the well ($V_0 = 160$ eV), since E must be less than V_0 for a particle trapped in a finite rectangular well. It's also instructive to compare these energies to the allowed energies of the infinite rectangular well, given in the previous section as

$$E_n = \frac{k_n^2 \hbar^2}{2m} = \frac{n^2 \pi^2 \hbar^2}{2ma^2}. \tag{5.7}$$

Inserting $m = 9.11 \times 10^{-31}$ kg and $a = 2.5 \times 10^{-10}$ m into this equation gives the lowest six energy levels ($n = 1$ to $n = 6$):

$E_1^\infty = 6.02$ eV	$E_3^\infty = 54.2$ eV	$E_5^\infty = 150.4$ eV
$E_2^\infty = 24.1$ eV	$E_4^\infty = 96.3$ eV	$E_6^\infty = 216.6$ eV

in which the ∞ superscript is a reminder that these energy levels pertain to an infinite rectangular well.

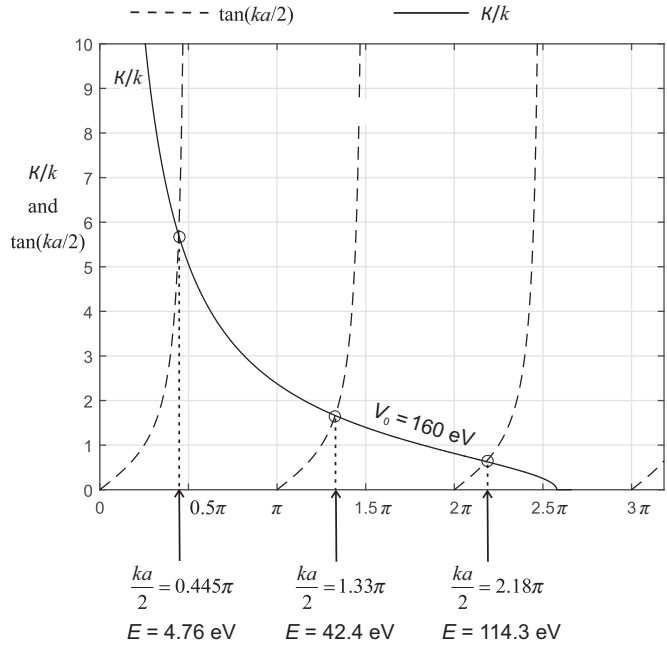


Figure 5.13 Solution values for $ka/2$ and E for 160-eV finite potential well (even case).

Remember that up to this point you've found only the energy levels corresponding to the even solutions of the Schrödinger equation for the finite rectangular well. As you'll later see, the process for finding the odd-parity solutions is almost identical to the process for finding the even-parity solutions, but before getting to that, you can compare the finite-well values to the first, third, and fifth energy levels of the infinite potential well. Since the lowest energy level of the finite well comes from the first even solution, after which the levels alternate between odd and even solutions, the even-solution energy levels of the finite rectangular well correspond to the odd-numbered ($n = 1, 3, 5 \ldots$) energy levels of the infinite rectangular well.

Thus the ground-state energy for this finite potential well ($E = 4.8$ eV) compares to the $n = 1$ energy level of $E_1^\infty = 6.02$ eV for the infinite well, which means that the finite-well ground-state energy is smaller than the infinite-well ground-state energy; the ratio is $4.8/6.02 = 0.8$. Comparing the next two even-solution energy levels for the finite well to E_3^∞ and E_5^∞ for the infinite well gives ratios of $42.4/54.2 = 0.78$ and $114.3/150.4 = 0.76$.

You can understand the reason that the energy levels of a finite well are smaller than those of a corresponding infinite well by comparing the finite-well wavefunction shown in Fig. 5.10 to the infinite-well wavefunctions shown in Fig. 5.2 (you can see more finite-well wavefunctions by looking ahead to Fig. 5.20). As described in Section 5.1, in the infinite-well case the wavefunctions must have zero amplitude at the edges of the well in order to match the zero-amplitude wavefunctions outside the well. But in the finite-well case the wavefunctions can have nonzero values at the edges of the well, where they must match up to the exponentially decaying wavefunctions in the evanescent region. That means that the finite-well wavefunctions can have longer wavelength than the corresponding infinite-well wavefunctions, and longer wavelength means smaller wavenumber k and smaller energy E. So it makes sense that the energy levels for a specified particle in a finite potential well are smaller than those of the same particle in an equal-width infinite well.

When you're considering the energy-level differences between a finite potential well and the corresponding infinite well, you should be careful not to lose sight of another important difference: a finite potential well has a finite number of allowed energy levels while an infinite well has an infinite number of allowed energy levels. As you'll see later in this section, every finite potential well has at least one allowed energy level, and the total number of allowed energies depends on both the depth and the width of the well.

You've seen the effect of varying the well depth on the number of allowed energies in Fig. 5.12, in which the number of even–parity solutions is one for the 2-eV well, two for the 60-eV well, and three for the 160-eV well. Those three wells have different depths, although they all share the same width ($a = 2.5 \times 10^{-10}$ m). But you may be wondering about the effect of varying the width of a finite potential well with a given depth on the number of allowed energies.

That effect is visible in Fig. 5.14, which shows the even-parity solutions for three finite wells. The depth of all three of these wells is 2 eV, but the widths of the wells are $a = 2.5 \times 10^{-10}$ m, $a = 10 \times 10^{-10}$ m, and $a = 25 \times 10^{-10}$ m, respectively.

As you can see in the figure, for finite rectangular wells of the same depth, wider wells may support a larger number of allowed energy levels. But the caveat discussed previously for increasing well depth also applies to increasing width; the increased width increases the number of allowed energy levels only if additional intersections of the κ/k curve and the even-solution $\tan(ka/2)$ or odd-solution $-\cot(ka/2)$ curve are produced.

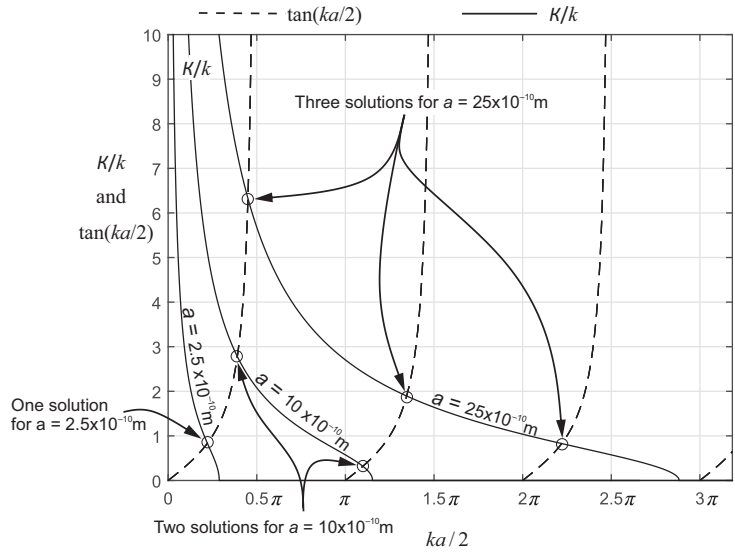


Figure 5.14 Effect of varying width of a finite rectangular well with $V_0 = 2$ eV.

The odd-solution version of the finite potential well transcendental equation is discussed presently, but before getting to that, you should consider an alternative form of this equation that's used in several quantum textbooks. That alternative form comes about by multiplying both sides of Eq. 5.33 by the factor $ka/2$:

$$\frac{ka}{2}\frac{\kappa}{k} = \frac{ka}{2}\tan\left(\frac{ka}{2}\right)$$

or

$$\frac{a}{2}\kappa = \frac{ka}{2}\tan\left(\frac{ka}{2}\right). \tag{5.37}$$

The effect of this multiplicative factor can be seen in Fig. 5.15. The curves representing the left-side function $\frac{a}{2}\kappa$ are circles centered on the origin, and the curves of the right-side function $\frac{ka}{2}\tan\left(\frac{ka}{2}\right)$ are scaled versions of those of $\tan(ka/2)$ of the original equation (Eq. 5.33).

To understand why the $\frac{a}{2}\kappa$ function produces circles when plotted with $ka/2$ on the horizontal axis, recall that the wavenumber k and the total energy E are related by the equations

$$k \equiv \sqrt{\frac{2m}{\hbar^2}E}. \tag{5.24}$$

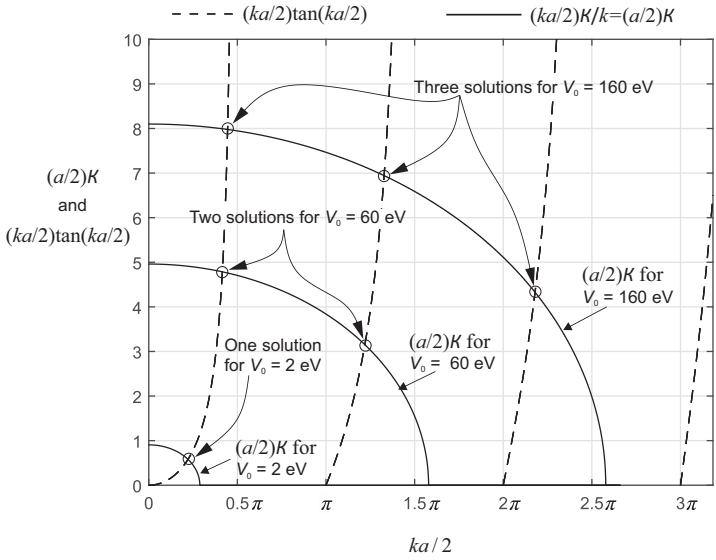


Figure 5.15 Finite potential well alternative graphical solution (even case).

and

$$E = 2\frac{\hbar^2\left(\frac{ka}{2}\right)^2}{ma^2}. \tag{5.36}$$

Now define a reference wavenumber k_0 as the wavenumber that the particle under consideration would have if its total energy were V_0 (in other words, if the particle's energy is just at the top of the finite potential well):

$$k_0 \equiv \sqrt{\frac{2m}{\hbar^2}V_0}, \tag{5.38}$$

which means that

$$V_0 = \frac{\hbar^2 k_0^2}{2m} = 2\frac{\hbar^2\left(\frac{k_0 a}{2}\right)^2}{ma^2}. \tag{5.39}$$

Also recall that κ is defined by the equation

$$\kappa \equiv \sqrt{\frac{2m}{\hbar^2}(V_0 - E)}, \tag{5.27}$$

so plugging in the expressions for E and V_0 from Eqs. 5.36 and 5.39 gives

$$\kappa = \sqrt{\frac{2m}{\hbar^2}\left(2\frac{\hbar^2\left(\frac{k_0 a}{2}\right)^2}{ma^2} - 2\frac{\hbar^2\left(\frac{ka}{2}\right)^2}{ma^2}\right)} = \sqrt{\frac{4}{a^2}\left[\left(\frac{k_0 a}{2}\right)^2 - \left(\frac{ka}{2}\right)^2\right]}$$

or

$$\frac{a}{2}\kappa = \sqrt{\left(\frac{k_0 a}{2}\right)^2 - \left(\frac{ka}{2}\right)^2}. \tag{5.40}$$

The left side of this equation is the left side of the modified form of the transcendental equation (Eq. 5.37), and this equation has the form of a circle of radius R:

$$x^2 + y^2 = R^2$$
$$y = \sqrt{R^2 - x^2}.$$

So plotting $\frac{a}{2}\kappa$ on the y-axis and $\frac{ka}{2}$ on the x-axis results in circles of radius $\frac{k_0 a}{2}$, as seen in Fig. 5.15.

If you compare the intersections of the curves in Fig. 5.15 with those of Fig. 5.12, you'll see that the $ka/2$ values of the intersections, and thus the allowed wavenumbers k and energies E, are the same. That's comforting, but it does raise the question of why you should bother with this alternative form of the transcendental equation. The answer is that in presenting the solutions for the finite potential well, several popular quantum texts use a substitution of variables that is a bit more understandable using this modified form of the transcendental equation. That substitution of variables is explained later in the chapter, so you can decide for yourself which form is more helpful.

The process for finding the allowed energy levels for the odd-parity solutions to the Schrödinger equation for the finite potential well closely parallels the approach used earlier in this section to find the allowed energies for the even-parity solutions.

Just as in the even-solution case, start by writing the continuity of the wavefunction amplitude at the left edge of the well ($x = -a/2$):

$$Ce^{\kappa\left(-\frac{a}{2}\right)} = B\sin\left[k\left(-\frac{a}{2}\right)\right], \tag{5.41}$$

and do the same for the slope of the wavefunction by equating the first spatial derivatives:

$$\kappa Ce^{\kappa\left(-\frac{a}{2}\right)} = kB\cos\left[k\left(-\frac{a}{2}\right)\right]. \tag{5.42}$$

As in the even-solution case, divide the continuous-spatial-derivative equation (Eq. 5.42) by the continuous-wavefunction equation (Eq. 5.41), which gives

$$\frac{\kappa C e^{\kappa\left(-\frac{a}{2}\right)}}{C e^{\kappa\left(-\frac{a}{2}\right)}} = \frac{kB\cos\left[k\left(-\frac{a}{2}\right)\right]}{B\sin\left[k\left(-\frac{a}{2}\right)\right]}, \tag{5.43}$$

or

$$\kappa = k\cot\left(-\frac{ka}{2}\right) = -k\cot\left(\frac{ka}{2}\right). \tag{5.44}$$

Dividing both sides of this equation by k gives

$$\frac{\kappa}{k} = -\cot\left(\frac{ka}{2}\right), \tag{5.45}$$

which is the odd-solution version of the transcendental equation (Eq. 5.33). Note that the left side of this equation is identical to the even-solution case, but the right side involves the negative cotangent rather than the positive tangent of $ka/2$.

The graphical approach to solving Eq. 5.45 is shown in Fig. 5.16. As you can see, the dashed lines representing the negative-cotangent function

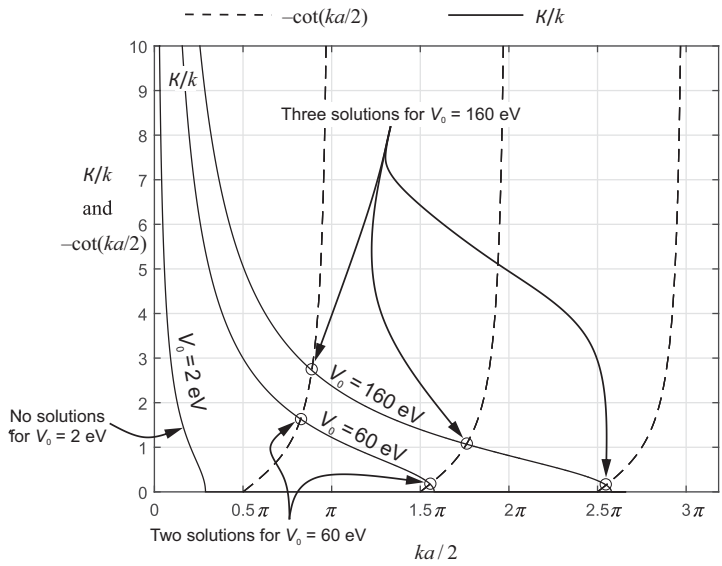


Figure 5.16 Finite potential well graphical solution (odd case) for three values of V_0.

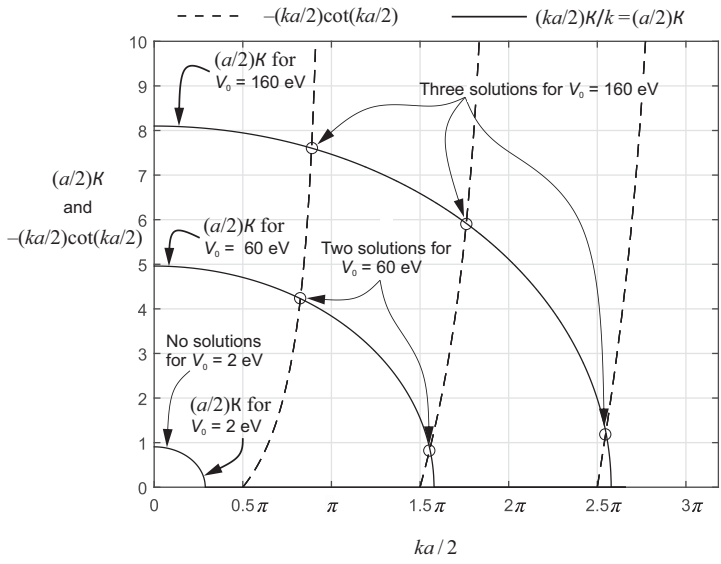


Figure 5.17 Finite potential well alternative graphical solution (odd case).

in this case are shifted along the horizontal axis by $\pi/2$, relative to the lines representing the positive-tangent function in the even-solution case.

An alternative form of the transcendental equation can also be found for the odd-parity solutions, as shown in Fig. 5.17. As expected, the allowed wavenumbers and energy levels are identical to those found using the original form of the transcendental equation.

One striking difference between Fig. 5.16 or Fig. 5.17 for odd-parity solutions and Fig. 5.12 or Fig. 5.15 for even-parity solutions is the lack of odd-parity solutions for the $V_0 = 2$ eV potential well. That's one consequence of the shifting of the negative-cotangent curves by $\pi/2$ to the right relative to the tangent curves of the even-parity case. In the even-parity case, the curve for $ka/2 = 0$ to $\pi/2$ begins at the origin and extends up and to the right, so, no matter how shallow and narrow the well is (that is, no matter how small V_0 and a are), its κ/k curve must cross the tan $(ka/2)$ curve. So every finite rectangular well is guaranteed to support at least one even-parity solution. But in the odd-parity case, the $-\cot(ka/2)$ curve crosses the horizontal axis at $ka/2 = 0.5\pi$, which means it's possible for the κ/k curve of a shallow (small V_0) and narrow (small a) potential well to get to $\kappa = 0$ (which occurs when $E = V_0$) without ever crossing one of the negative cotangent curves.

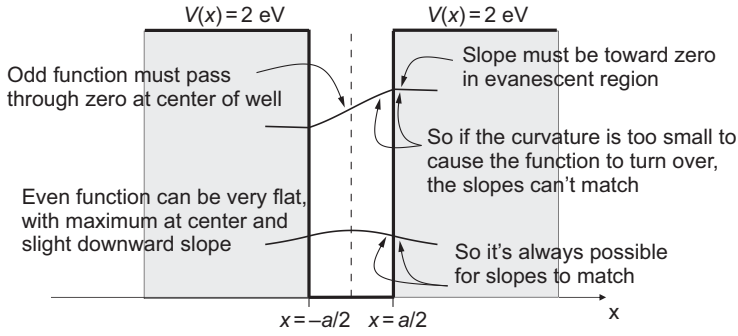

V(x) = 2 eV
V(x) = 2 eV
Odd function must pass through zero at center of well
Slope must be toward zero in evanescent region
So if the curvature is too small to cause the function to turn over, the slopes can't match
Even function can be very flat, with maximum at center and slight downward slope
So it's always possible for slopes to match
x = –a/2
x = a/2
x

Figure 5.18 A narrow, shallow potential well supporting only one even solution.

So what's happening physically that ensures at least one even-parity solution but allows no odd-parity solutions for sufficiently shallow and narrow potential wells? To understand that, consider the lowest-energy (and therefore lowest-curvature) even wavefunction. That ground-state wavefunction may be almost flat as it extends from the left edge to the right edge of the well, with very small slope everywhere, as indicated by the even-function curve in Fig. 5.18. That small slope must be matched at the edges of the well by the decaying-exponential function in the evanescent region, which means that the decay constant κ must have a small value. And since κ is proportional to the square root of $V_0 - E$, it's always possible to find an energy E sufficiently close in value to V_0 to cause the spatial decay rate (set by the value of κ) to match the small slope of the wavefunction at the inside of each edge of the well.

The situation is very different for odd-parity solutions, as indicated by the odd-function curve in Fig. 5.18. This well has a depth of only 2 eV and width of 2.5×10^{-10} m, which means that any particle trapped in this well must have energy E no greater than 2 eV, so the curvature will be small. But the small curvature due to the low value of E means that the odd-parity wavefunction does not have room to "turn over" in the space between the center of the well (through which all odd wavefunctions must cross) and the edge of the well. So there's no hope of matching the slope of the oscillating wavefunction within the well to the slope of the decaying wavefunction in the evanescent region.

Hence a 2-eV finite rectangular well of width $a = 2.5 \times 10^{-10}$ m can support one even solution but no odd solutions. If, however, you increase the well width, odd solutions become possible even for a shallow well, as you can see in Fig. 5.19.

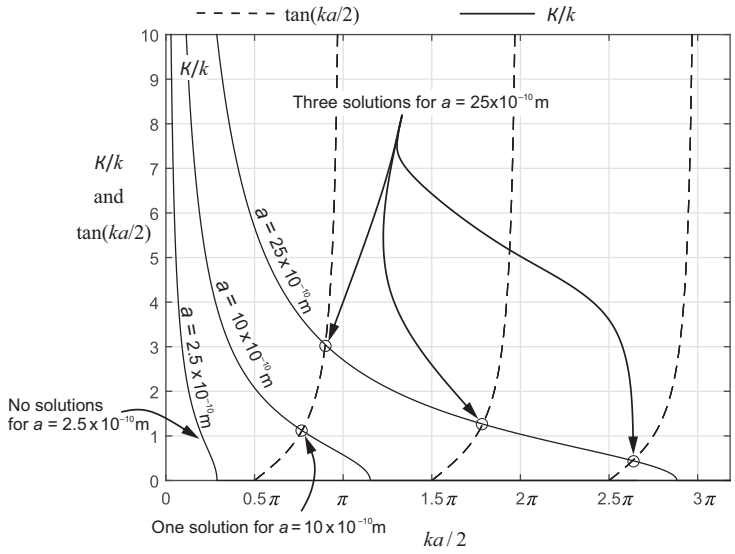


Figure 5.19 Effect of varying width of finite rectangular well for odd solutions with $V_0 = 2$ eV.

You can think of this lack of an odd-parity solution as the κ/k curve not intersecting the negative cotangent curve, or as the slope of the odd wavefunction at the inside edge of the well not matching the slope of the exponentially decaying wavefunction at the outside edge of the well. Either way, the conclusion is that "small" or "weak" potential wells (that is, shallow or narrow finite potential wells) may not support any odd-parity solutions, but they always support at least one even-parity solution.

To determine the allowed wavenumbers and energies of each of the three potential wells considered earlier (with potential energies $V_0 = 2$ eV, 60 eV, or 160 eV), you can use either Fig. 5.16 or 5.17. In this case, the intersections occur at $ka/2$ values of 0.888π, 1.76π, and 2.55π. Plugging those values into Eq. 5.36 gives the allowed energy values of 19.0 eV, 74.6 eV, and 156.3 eV.

As mentioned previously, the corresponding energy levels of the same particle in an infinite rectangular well with the same width are $E_2^\infty = 24.1$ eV, $E_4^\infty = 96.3$ eV, and $E_6^\infty = 216.6$ eV. So as for the even solutions, the energy levels of the 160-eV finite well are 70% to 80% of the energies of the corresponding infinite well.

The wavefunctions for all six allowed energy levels for a finite potential well with $V_0 = 160$ eV and width $a = 2.5 \times 10^{-10}$ m are shown in Fig. 5.20.

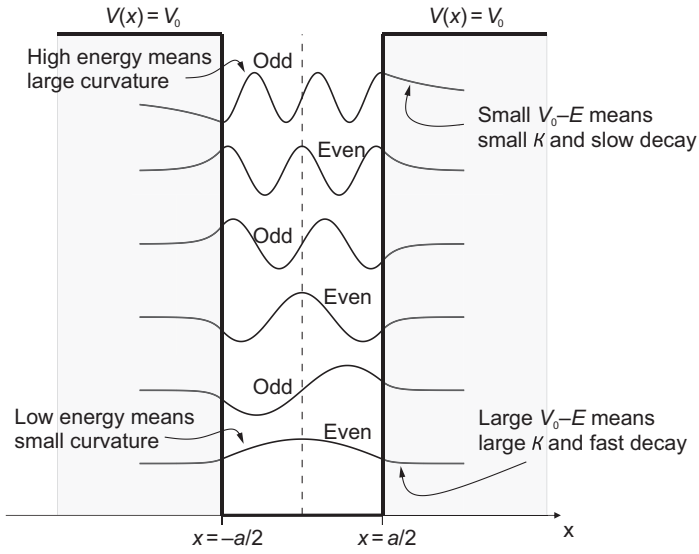


Figure 5.20 Alternating even and odd solutions for finite potential well.

As you can see, the ground-state wavefunction is an even function with low curvature (due to small value of E, which means small wavenumber k) and fast decay in the evanescent region (due to large value of $V_0 - E$, which means large decay constant κ). The wavefunctions for the five allowed excited states alternate between odd and even, and as energy E and wavenumber k increase, the curvature becomes larger, meaning that more cycles fit within the well. But larger energy E means smaller values of $V_0 - E$, so the decay constant κ decreases, and smaller decay rates mean larger penetration into the classically forbidden region.

The last bit of business for the finite potential well is the substitution of variables mentioned earlier in this section. It's worth some of your time to understand this process if you're planning to read a comprehensive text on quantum mechanics, because this substitution of variables, or some variant of it, is commonplace. But if you've worked through the material in this section, this substitution shouldn't cause difficulty, because it involves quantities that have played an important role in the previous discussion.

The principal substitution is this: make a new variable z, defined as the product of the wavenumber k and the half-width of the potential well $a/2$, so $z \equiv ka/2$. And exactly what does z represent? Since the wavenumber k is related to the wavelength λ through the equation $k = 2\pi/\lambda$, the product

$ka/2$ represents the number of wavelengths in the half-width $a/2$, converted to radians by the factor of 2π. For example, if the half-width of the potential well $a/2$ is equal to one wavelength, then $z = ka/2$ has a value of 2π radians, and if $a/2$ equals two wavelengths, then $z = ka/2$ has a value of 4π radians. So z in radians is proportional to the width of the well in wavelengths.

It's also useful to understand the relationship between z and the total energy E. Using the relationship between wavenumber k and total energy E (Eq. 5.24), you can write the variable z in terms of energy as

$$z \equiv k\frac{a}{2} = \left(\sqrt{\frac{2m}{\hbar^2}E}\right)\frac{a}{2} \tag{5.46}$$

or, solving for E,

$$E = \left(\frac{2}{a}\right)^2\left(\frac{\hbar^2}{2m}\right)z^2. \tag{5.47}$$

Inserting the expression for z into the even-solution transcendental equation $\kappa/k = \tan(ka/2)$ (Eq. 5.33) gives

$$\frac{\kappa}{\frac{2}{a}z} = \tan(z). \tag{5.48}$$

This doesn't appear to be much of an improvement, but the advantage of using z becomes clear if you also do a similar substitution of variables for κ. To do that, begin by defining a variable z_0 as the product of the reference wavenumber k_0 and the well half-width:

$$z_0 \equiv k_0 a/2. \tag{5.49}$$

Recall that the reference wavenumber k_0 is defined by Eq. 5.38 as the wavenumber of a quantum particle with energy E equal to the depth of the finite potential well V_0. That means that z_0 can be written in terms of well depth V_0:

$$z_0 = \frac{k_0 a}{2} = \sqrt{\frac{2m}{\hbar^2}V_0}\frac{a}{2} \tag{5.50}$$

or, solving for V_0,

$$V_0 = \left(\frac{2}{a}\right)^2\left(\frac{\hbar^2}{2m}\right)z_0^2. \tag{5.51}$$

Now insert the expressions for E and V_0 (Eqs. 5.47 and 5.51) into the definition of κ (Eq. 5.27):

$$\kappa = \sqrt{\frac{2m}{\hbar^2}(V_0 - E)} = \sqrt{\frac{2m}{\hbar^2}\left[\left(\frac{2}{a}\right)^2\left(\frac{\hbar^2}{2m}\right)z_0^2 - \left(\frac{2}{a}\right)^2\left(\frac{\hbar^2}{2m}\right)z^2\right]}$$
$$= \sqrt{\frac{4}{a^2}(z_0^2 - z^2)} = \frac{2}{a}\sqrt{z_0^2 - z^2}.$$

The payoff for all this work comes from substituting this expression for κ into Eq. 5.48. That substitution gives

$$\frac{\kappa}{\frac{2}{a}z} = \frac{\frac{2}{a}\sqrt{z_0^2 - z^2}}{\frac{2}{a}z} = \tan(z)$$

or

$$\sqrt{\frac{z_0^2}{z^2} - 1} = \tan(z). \tag{5.52}$$

This form of the even-solution transcendental equation is entirely equivalent to Eq. 5.33, and it's one of the versions that you're likely to encounter in other quantum texts. When dealing with this equation, just remember that z is a measure of the total energy of the particle ($z \propto \sqrt{E}$), and z_0 is related to the depth of the well ($z_0 \propto \sqrt{V_0}$). So for a given mass m and well width a, higher energy means larger z, and deeper well means larger z_0.

The process of solving this equation graphically is identical to the process described earlier, and you can see an example of that graphical solution for three values of z_0 in Fig. 5.21. In this plot, the curves representing $\sqrt{\frac{z_0^2}{z^2} - 1}$ have the same shape as the κ/k curves in Fig. 5.12, and the dashed curves representing $\tan(z)$ are identical to the $\tan(ka/2)$ curves, since $z = ka/2$.

If you're wondering why the values $z_0 = 1$, 5, and 8 were chosen for Fig. 5.21, consider the result of converting these values of z_0 into the well depth V_0, assuming that the mass m and well width a are the same as those used in Fig. 5.12. For $z_0 = 1$, Eq. 5.51 tells you that V_0 is

$$V_0 = \left(\frac{2}{a}\right)^2\left(\frac{\hbar^2}{2m}\right)z_0^2 = \left(\frac{2}{2.5 \times 10^{-10}\text{m}}\right)^2\left(\frac{(1.06 \times 10^{-34}\text{JS})^2}{2(9.11 \times 10^{-31}\text{kg})}\right)(1)^2$$
$$= 3.91 \times 10^{-19}\ \text{J} = 2.4\ \text{eV}.$$

Doing the same calculation for $z_0 = 5$ and $z_0 = 8$ reveals that $z_0 = 5$ corresponds to $V_0 = 61.0$ eV and $z_0 = 8$ corresponds to $V_0 = 156.1$ eV for

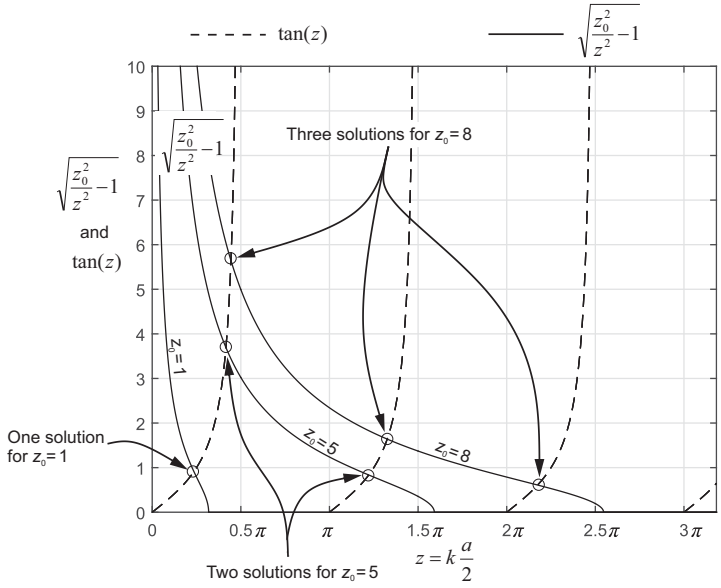


Figure 5.21 Finite well even-parity graphical solution using z-substitution.

the given values of m and a. So the equations $z_0 = 1$, 5, and 8 correspond to well depths close to the values of 2 eV, 60 eV, and 160 eV used in Fig. 5.12. This is not to imply that z_0 is restricted to integer values; choosing $z_0 = 0.906$, 4.96, and 8.10 makes the V_0 values 2.0, 60.0, and 160.0 eV.

An alternative form of this equation, equivalent to Eq. 5.37, can be found quite easily using the variable substitution $z = ka/2$ and $z_0 = k_0 a/2$. For that version, multiply both sides of Eq. 5.52 by z:

$$z\sqrt{\frac{z_0^2}{z^2} - 1} = z\tan(z)$$

or

$$\sqrt{z_0^2 - z^2} = z\tan(z). \tag{5.53}$$

This is the "z-version" of Eq. 5.37, and the simplicity of getting this result demonstrates one of the advantages of using this substitution of variables. Another benefit is that the form of this equation makes clear the circular nature of the curves produced by plotting its left side on the vertical axis with the horizontal axis representing z.

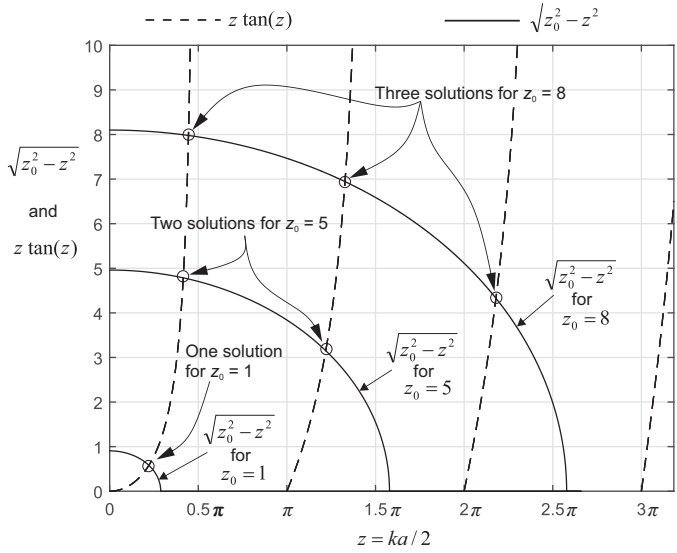


Figure 5.22 Finite potential well alternative graphical solution (even case) using z-substitution.

You can see those curves in Fig. 5.22, using the same parameters (m and a) used previously. As before, three values of well depth are used in this plot, corresponding to $z_0 = 1$, 5, and 8.

Careful comparison of Figs. 5.21 and 5.22 shows that the values of $z = ka/2$ at which the curves representing the left and right sides of the even-solution transcendental equation intersect are the same, so you should feel free to use whichever version you prefer.

As you've probably anticipated, the same substitution of variables $z = ka/2$ and $z_0 = k_0 a/2$ can be applied to the odd-parity solutions for the finite potential well. Recall that the transcendental equation for the odd solutions has $-\cot(ka/2)$ rather than $\tan(ka/2)$ on the right side, and making the z and z_0 substitutions into Eq. 5.45 gives

$$\sqrt{\frac{z_0^2}{z^2} - 1} = -\cot(z), \tag{5.54}$$

for which the graphical solution is shown in Fig. 5.23. As expected, the three curves representing the left side of this equation for z_0 values of 1, 5, and 8 are identical to the corresponding even-solution curves, but the negative cotangent

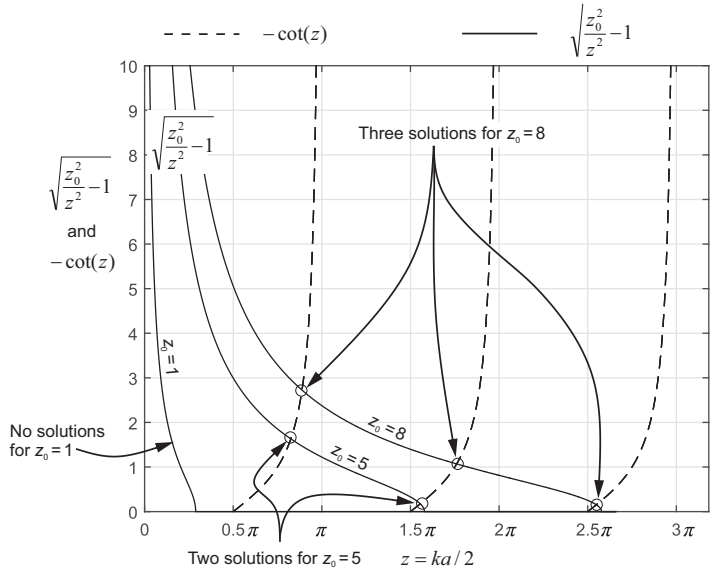


Figure 5.23 Finite potential well graphical solution (odd case) using z-substitution.

curves are offset along the horizontal (z) axis by $\pi/2$ relative to the even-solution case.

Multiplying both sides by z gives the alternative equation:

$$\sqrt{z_0^2 - z^2} = -z\cot(z), \tag{5.55}$$

for which the graphical solutions are shown in Fig. 5.24.

It's fair to say that the process of finding the allowed wavefunctions and energy levels of the finite potential well has proven to be somewhat more complicated than the equivalent process for the infinite well. The payback for the extra effort required by that process is that the finite well is a more realistic representation of physically realizable conditions than the infinite well. But the use of piecewise-constant potentials, which means zero force everywhere except at the edges of the well, limits the applicability of the finite-well model. In the final section of this chapter, you can work through an example of a potential well in which the potential is not constant (meaning the force is nonzero) within the well. That example is called the quantum harmonic oscillator.

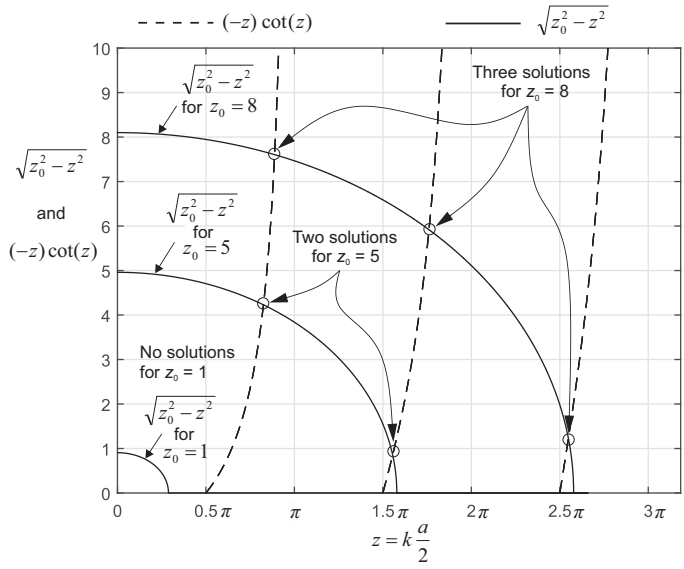


Figure 5.24 Finite potential well alternative graphical solution (odd case) using z-substitution.

5.3 Harmonic Oscillator

The quantum **harmonic oscillator** is worth your attention for several reasons. One of those reasons is that it provides a instructive example of the application of several of the concepts of previous sections and chapters. But in addition to applying concepts you've seen before, in finding the solutions to the quantum harmonic oscillator problem, you'll also see how to use several techniques that were not required for problems such as the infinite and finite rectangular well. Equally important is the usefulness of these techniques for other problems, because the potential-energy function $V(x)$ of the harmonic oscillator is a reasonable approximation for other potential energy functions in the vicinity of a potential minimum. This means that the harmonic oscillator, although idealized in this treatment, has a strong connection to several real-world configurations.

If it's been a while since you looked at the classical harmonic oscillator, you may want to spend some time reviewing the basics of the behavior of a system such as a mass sliding on a frictionless horizontal surface while attached to a spring. In the classical case, this type of system oscillates with constant

total energy, continuously exchanging potential and kinetic energy as it moves from the equilibrium position to the "turning points" at which its direction of motion reverses. The potential energy of that object is zero at the equilibrium position and maximum at the turning points at which the spring is maximally compressed or extended. Conversely, the kinetic energy is maximum as the object passes through equilibrium and zero when the object's velocity passes through zero at the turning points. The object moves fastest at equilibrium and slowest at the turning points, which means that measurements of position taken at random times are more likely to yield results near the turning points, because the object spends more time there.

As you'll see in this section, the behavior of the quantum harmonic oscillator is quite different from its classical counterpart, but several aspects of the classical harmonic oscillator are relevant to the quantum case. One of those aspects is the quadratic form of the potential energy, usually written as

$$V(x) = \frac{1}{2}kx^2, \tag{5.56}$$

in which x represents the distance of the object from the equilibrium position and k represents the "spring constant" (the force on the object per unit distance from the equilibrium position). This quadratic relationship between potential energy and position pertains to any restoring force that increases linearly with distance, that is, any force that obeys **Hooke's Law**:

$$F = -kx, \tag{5.57}$$

in which the minus sign indicates that the force is always in the direction toward the equilibrium point (opposite to the direction of displacement from equilibrium). You can see the relationship between Hooke's Law and quadratic potential energy by writing force as the negative gradient of the potential energy:

$$F = -\frac{\partial V}{\partial x} = -\frac{\partial\left(\frac{1}{2}kx^2\right)}{\partial x} = -\frac{2kx}{2} = -kx. \tag{5.58}$$

Another useful result from the classical harmonic oscillator is that the motion of the object is sinusoidal, with angular frequency ω give by

$$\omega = \sqrt{\frac{k}{m}}, \tag{5.59}$$

in which k represents the spring constant and m represents the mass of the object.

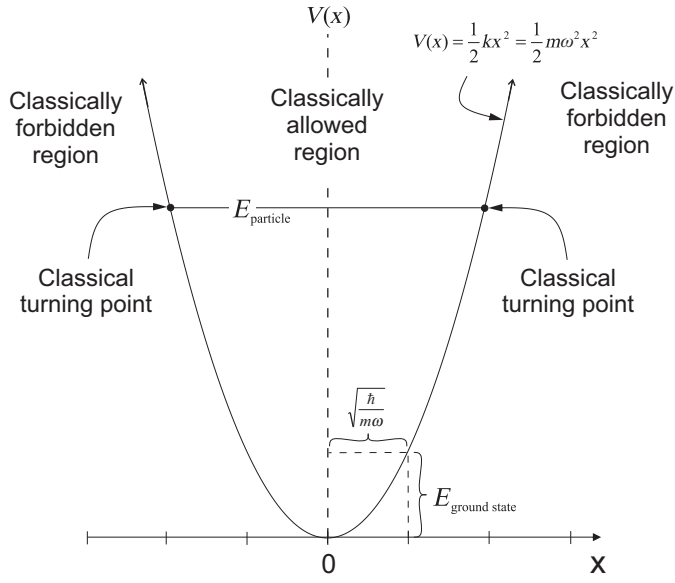


Figure 5.25 Harmonic oscillator potential.

You can see a plot of the potential energy of a harmonic oscillator as a function of distance x from equilibrium in Fig. 5.25 (it's the parabolic curve – the other aspects of the figure are explained shortly). Notice that the potential becomes infinitely large as $x \to \pm\infty$. As you saw in the case of the infinite rectangular well, the amplitude of the wavefunction $\psi(x)$ must be zero in regions in which the potential energy is infinite; this provides the boundary conditions for the wavefunctions of the quantum oscillator.

As in the potential wells of the previous sections, you can find the energy levels and wavefunctions of the quantum harmonic oscillator by using separation of variables and solving the TISE (Eq. 3.40). For the quantum harmonic oscillator, that equation looks like this:

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} + \frac{1}{2}kx^2\psi(x) = E\psi(x). \tag{5.60}$$

In quantum mechanics, it's customary to write equations and solutions in terms of angular frequency ω rather than spring constant k. Solving Eq. 5.59 for k gives $k = m\omega^2$, and plugging that into the time-independent Schrödinger equation gives

$$\frac{d^2\psi(x)}{dx^2} - \frac{2m}{\hbar^2}\left[\frac{1}{2}m\omega^2 x^2\psi(x)\right] = -\frac{2m}{\hbar^2}E\psi(x)$$
$$\frac{d^2\psi(x)}{dx^2} - \left[\frac{m^2\omega^2}{\hbar^2}x^2\psi(x)\right] + \frac{2m}{\hbar^2}E\psi(x) = 0$$
$$\frac{d^2\psi(x)}{dx^2} + \left[\frac{2m}{\hbar^2}E - \frac{m^2\omega^2}{\hbar^2}x^2\right]\psi(x) = 0. \tag{5.61}$$

This version of the Schrödinger equation is considerably more difficult to solve than the version used for the infinite rectangular well in Section 5.1 and the finite rectangular well in Section 5.2, and that's because of the x^2 in the potential term (recall that the potential energy $V(x)$ was taken as constant over each of the regions in those sections). Those piecewise-constant potentials led to constant wavenumber k (proportional to $\sqrt{E}$, the distance above the bottom of the well) inside the well and decay constant κ (proportional to $\sqrt{V_0 - E}$, the distance below the top of the well) outside the well. But in this case the depth of the well varies continuously with x, so a different approach is required.

If you've looked at the harmonic-oscillator material in comprehensive quantum texts, you may have noticed that there are two different approaches to finding the energy levels and wavefunctions of the quantum harmonic oscillator, sometimes called the "analytic" approach and the "algebraic" approach. The analytic approach uses a power series to solve Eq. 5.61, and the algebraic approach involves factoring Eq. 5.61 and using a type of operator called a **"ladder" operator** to determine allowed energy levels and wavefunctions. In keeping with this book's goal of preparing you for future encounters with the literature of quantum mechanics, you'll find the basics of both of these approaches in this section.

Even if you've had limited exposure to differential equations, the analytic power-series approach to solving the TISE for the harmonic oscillator is reasonably comprehensible, and once you've seen how it works, you should be happy to add this technique to your toolbox.

The bookkeeping is a bit less tedious if you make two variable substitutions before starting down the analytic path. Both of those substitutions are motivated by the same idea, which is to replace a dimensional variable, such as the energy E and position x, with a dimensionless quantity. In each case, you can think of this as dividing the quantity by a reference quantity, such as E_{ref} and x_{ref}. In this section, the dimensionless version of energy is called ϵ, defined like this:

$$\epsilon \equiv \frac{E}{E_{ref}} = \frac{E}{\left(\frac{1}{2}\hbar\omega\right)}, \tag{5.62}$$

in which the reference energy is $E_{ref} = \hbar\omega/2$. You can easily verify that this expression for E_{ref} has dimensions of energy, but where does that factor of $1/2$ come from, and what's ω?

The answers to those questions will become clear once you've seen the energy levels for the harmonic oscillator, but the short version is that ω is the angular frequency of the ground-state (lowest-energy) wavefunction, and $\hbar\omega/2$ turns out to be the ground-state energy of the quantum harmonic oscillator.

The dimensionless version of position is called ξ, defined by

$$\xi = \frac{x}{x_{ref}} = \frac{x}{\sqrt{\frac{\hbar}{m\omega}}}, \tag{5.63}$$

in which the reference position is $x_{ref} = \sqrt{\frac{\hbar}{m\omega}}$. As always, it's a good idea to check that this expression for x_{ref} has dimensions of position.

So what does $\sqrt{\frac{\hbar}{m\omega}}$ represent? As in the case of E_{ref}, the answer will be clear once you've determined the energy levels of the quantum harmonic oscillator, but here's a preview: $\sqrt{\frac{\hbar}{m\omega}}$ is the distance to the classical turning point of a harmonic oscillator for a particle in the ground state. As you'll see in this section, quantum particles don't behave like classical harmonic oscillators, but the distance to the classical turning point is nonetheless a convenient reference. Both E_{ref} and x_{ref} are shown in Fig. 5.25.

To get these dimensionless quantities into Eq. 5.61, you can't simply divide the energy term by E_{ref} and the position term by x_{ref}. Instead, start by solving Eqs. 5.62 and 5.63 for E and x, respectively:

$$E = \epsilon E_{ref} = \epsilon\left(\frac{1}{2}\hbar\omega\right) \tag{5.64}$$

and

$$x = \xi x_{ref} = \xi\left(\sqrt{\frac{\hbar}{m\omega}}\right). \tag{5.65}$$

Next, it's necessary to work on the second-order spatial derivative d^2/dx^2. Taking the first spatial derivative of x with respect to ξ gives

$$\begin{aligned} \frac{dx}{d\xi} &= \sqrt{\frac{\hbar}{m\omega}}, \\ dx &= \sqrt{\frac{\hbar}{m\omega}}d\xi, \end{aligned} \tag{5.66}$$

and

$$dx^2 = \frac{\hbar}{m\omega} d\xi^2. \tag{5.67}$$

Now plug in these expressions for E, x, and dx^2 into Eq. 5.61, which gives

$$\begin{aligned}
\frac{d^2\psi(\xi)}{\frac{\hbar}{m\omega}d\xi^2} + \left[\frac{2m}{\hbar^2}\epsilon\left(\frac{1}{2}\hbar\omega\right) - \frac{m^2\omega^2}{\hbar^2}\left(\xi\sqrt{\frac{\hbar}{m\omega}}\right)^2\right]\psi(\xi) &= 0 \\
\frac{m\omega}{\hbar}\frac{d^2\psi(\xi)}{d\xi^2} + \left(\frac{m\omega}{\hbar}\epsilon - \frac{m\omega}{\hbar}\xi^2\right)\psi(\xi) &= 0 \\
\frac{d^2\psi(\xi)}{d\xi^2} + \left(\epsilon - \xi^2\right)\psi(\xi) &= 0.
\end{aligned} \tag{5.68}$$

Differential equations of this type are called **Weber equations**, for which the solutions are known to be products of Gaussian functions and **Hermite polynomials**. Before seeing how that comes about, you should step back and consider what Eq. 5.68 is telling you.

If you've read the curvature discussion in Chapters 3 and 4, you know that the second spatial derivative $d^2\psi/dx^2$ represents the curvature of the wavefunction ψ over distance. From the definitions just given, you also know that ϵ is proportional to energy E and ξ^2 is proportional to the square of position x^2, so Eq. 5.68 means that the magnitude of the curvature of harmonic-oscillator wavefunctions increases as energy increases, but for a given energy, the wavefunction curvature decreases with distance from the center of the potential well.

That analysis gives you a general idea of the behavior of quantum oscillator wavefunctions, but the details of that behavior can only be determined by solving Eq. 5.68. To do that, it helps to consider what the equation tells you about the asymptotic behavior of the solutions $\psi(\xi)$ (that is, the behavior at very large or very small values of ξ). That's useful because you may be able to separate out the behavior of the solution in one regime from that in another, and the differential equation may be simpler to solve in those regimes.

It's not hard to see how that works with Eq. 5.68, which for large ξ (and hence large values of x) looks like this:

$$\begin{aligned}
\frac{d^2\psi(\xi)}{d\xi^2} - \xi^2\psi(\xi) &\approx 0 \\
\frac{d^2\psi(\xi)}{d\xi^2} &\approx \xi^2\psi(\xi),
\end{aligned} \tag{5.69}$$

in which the ϵ term is negligible relative to the ξ^2 term for large ξ.

The solutions to this equation for large ξ are

$$\psi(\xi \to \pm\infty) = Ae^{\frac{\xi^2}{2}} + Be^{-\frac{\xi^2}{2}}, \tag{5.70}$$

but for the harmonic oscillator, the potential energy $V(x)$ increases without limit as x (and therefore ξ) goes to positive or negative infinity. As mentioned previously, this means that the wavefunction $\psi(\xi)$ must go to zero as $\xi \to \pm\infty$. That rules out the positive-exponential solutions, so the coefficient A must be zero.

That leaves the negative-exponential term as the dominant portion of $\psi(\xi)$ at large positive and negative values of ξ, so you can write

$$\psi(\xi) = f(\xi)e^{-\frac{\xi^2}{2}}, \tag{5.71}$$

in which the $f(\xi)$ represents a function that determines the behavior of $\psi(\xi)$ at small values of ξ, and the constant coefficient B has been absorbed into the function $f(\xi)$.

What good has it done to separate out the asymptotic behavior of $\psi(\xi)$? To see that, look at what happens if you plug the expression for $\psi(\xi)$ given by Eq. 5.71 into Eq. 5.68:

$$\frac{d^2\left[f(\xi)e^{-\frac{\xi^2}{2}}\right]}{d\xi^2} + \left(\epsilon - \xi^2\right)f(\xi)e^{-\frac{\xi^2}{2}} = 0. \tag{5.72}$$

Now apply the product rule of differentiation to the first spatial derivative:

$$\begin{aligned}\frac{d\left[f(\xi)e^{-\frac{\xi^2}{2}}\right]}{d\xi} &= \frac{df(\xi)}{d\xi}e^{-\frac{\xi^2}{2}} + f(\xi)\frac{d\left(e^{-\frac{\xi^2}{2}}\right)}{d\xi} \\ &= \frac{df(\xi)}{d\xi}e^{-\frac{\xi^2}{2}} + f(\xi)\left(-\xi e^{-\frac{\xi^2}{2}}\right) \\ &= e^{-\frac{\xi^2}{2}}\left[\frac{df(\xi)}{d\xi} - \xi f(\xi)\right]\end{aligned}$$

and taking another spatial derivative gives

$$\begin{aligned}\frac{d^2\left[f(\xi)e^{-\frac{\xi^2}{2}}\right]}{d\xi^2} &= \frac{d\left\{e^{-\frac{\xi^2}{2}}\left[\frac{df(\xi)}{d\xi} - \xi f(\xi)\right]\right\}}{d\xi} \\ &= \frac{d\left(e^{-\frac{\xi^2}{2}}\right)}{d\xi}\left[\frac{df(\xi)}{d\xi} - \xi f(\xi)\right] + e^{-\frac{\xi^2}{2}}\frac{d}{d\xi}\left[\frac{df(\xi)}{d\xi} - \xi f(\xi)\right]\end{aligned}$$

$$
\begin{aligned}
&= -\xi e^{-\frac{\xi^2}{2}} \frac{df(\xi)}{d\xi} - \xi e^{-\frac{\xi^2}{2}} [-\xi f(\xi)] + e^{-\frac{\xi^2}{2}} \frac{d^2 f(\xi)}{d\xi^2} \\
&\quad + e^{-\frac{\xi^2}{2}} \left[-f(\xi) - \xi \frac{df(\xi)}{d\xi} \right] \\
&= e^{-\frac{\xi^2}{2}} \left[-\xi \frac{df(\xi)}{d\xi} + \xi^2 f(\xi) + \frac{d^2 f(\xi)}{d\xi^2} - f(\xi) - \xi \frac{df(\xi)}{d\xi} \right] \\
&= e^{-\frac{\xi^2}{2}} \left[\frac{d^2 f(\xi)}{d\xi^2} - 2\xi \frac{df(\xi)}{d\xi} + f(\xi)(\xi^2 - 1) \right].
\end{aligned}
$$

Plugging this into Eq. 5.72 gives

$$
e^{-\frac{\xi^2}{2}} \left[\frac{d^2 f(\xi)}{d\xi^2} - 2\xi \frac{df(\xi)}{d\xi} + f(\xi)(\xi^2 - 1) \right] + \left(\epsilon - \xi^2\right) f(\xi) e^{-\frac{\xi^2}{2}} = 0
$$

or

$$
e^{-\frac{\xi^2}{2}} \left[\frac{d^2 f(\xi)}{d\xi^2} - 2\xi \frac{df(\xi)}{d\xi} + f(\xi)(\epsilon - 1) \right] = 0. \tag{5.73}
$$

Since this equation must be true for all values of ξ, and the leading exponential factor cannot be zero everywhere, the term in square brackets must equal zero:

$$
\frac{d^2 f(\xi)}{d\xi^2} - 2\xi \frac{df(\xi)}{d\xi} + f(\xi)(\epsilon - 1) = 0. \tag{5.74}
$$

It may seem that all this work has simply gotten you to another second-order differential equation, but this one is amenable to solution by the power-series approach. To do that, write the function $f(\xi)$ as a power series in ξ:

$$
\begin{aligned}
f(\xi) &= a_0 + a_1 \xi + a_2 \xi^2 + \cdots \\
&= \sum_{n=0}^{\infty} a_n \xi^n.
\end{aligned}
$$

Note that for the quantum harmonic oscillator it's customary to start the index at $n = 0$ rather than $n = 1$, so the ground-state (lowest-energy) wavefunction will be called ψ_0, and the lowest energy level will be called E_0. Representing $f(\xi)$ with this power series makes the first and second spatial derivatives of $f(\xi)$

$$
\frac{df(\xi)}{d\xi} = \sum_{n=0}^{\infty} n a_n \xi^{n-1}
$$

and

$$\frac{d^2f(\xi)}{d\xi^2} = \sum_{n=0}^{\infty} n(n-1)a_n\xi^{n-2}.$$

Inserting these into Eq. 5.74 gives

$$\sum_{n=0}^{\infty} n(n-1)a_n\xi^{n-2} - 2\xi\sum_{n=0}^{\infty} na_n\xi^{n-1} + \sum_{n=0}^{\infty} a_n\xi^n(\epsilon-1) = 0. \tag{5.75}$$

An equation such as this can be made much more useful by grouping the terms that have the same power of ξ. That's because all of the terms with the same power of ξ must sum to zero. To understand why that's true, consider this: Eq. 5.75 says that all terms of all powers must sum to zero, but terms of one power cannot cancel terms of a different power (terms of different powers may cancel one another for a certain value of ξ, but not over all values of ξ). So if you group the terms of Eq. 5.75 that have the same power, you can be certain that the coefficients of those terms sum to zero.

And although it may seem like a chore to group the same-power terms of Eq. 5.75, the second and third summations already have the same powers of ξ. That power is n, since there's an additional factor of ξ in the second summation, and $(\xi)(\xi^{n-1}) = \xi^n$. Now look carefully at the first summation, for which the $n = 0$ and $n = 1$ terms both contribute nothing to the sum. That means you can simply renumber the indices by letting $n \to n+2$, which means that summation also contains ξ^n. So Eq. 5.75 may be written as

$$\sum_{n=0}^{\infty}(n+2)(n+1)a_{n+2}\xi^n - \sum_{n=0}^{\infty} 2na_n\xi^n + \sum_{n=0}^{\infty} a_n\xi^n(\epsilon-1) = 0$$

$$\sum_{n=0}^{\infty}\left[(n+2)(n+1)a_{n+2} - 2na_n + a_n(\epsilon-1)\right]\xi^n = 0,$$

which means that the coefficients of ξ^n for each value of n must sum to zero:

$$(n+2)(n+1)a_{n+2} - 2na_n + a_n(\epsilon-1) = 0$$

$$a_{n+2} = \frac{2n+(1-\epsilon)}{(n+2)(n+1)}a_n. \tag{5.76}$$

This is a recursion relation that relates any coefficient a_n to the coefficient a_{n+2} that is two steps higher. So if you know any one of the even coefficients, you can determine all the higher even components using this equation, and you can

find all the lower even coefficients (if any) by re-indexing this equation, letting n become $n-2$. Likewise, if you know any one of the odd coefficients, you can find all the other odd coefficients. For example, if you know the coefficient a_0, you can determine a_2, a_4, etc., and if you know a_1 you can determine a_3, a_5, and so on to infinity.

An issue arises, however, if you consider what this equation says about the ratio a_{n+2}/a_n for large n. That ratio is

$$\frac{a_{n+2}}{a_n} = \frac{2n + (1-\epsilon)}{(n+2)(n+1)}, \tag{5.77}$$

and for large values of n, the terms containing n dominate the other terms in both the numerator and the denominator. So this ratio converges to

$$\frac{a_{n+2}}{a_n} = \frac{2n + (1-\epsilon)}{(n+2)(n+1)} \xrightarrow[\text{large n}]{} \frac{2n}{(n)(n)} = \frac{2}{n}. \tag{5.78}$$

Why is this a problem? Because $2/n$ is exactly what the ratio of the even or odd terms in the power series for the function e^{ξ^2} converges to, and if the ratio a_{n+2}/a_n behaves like e^{ξ^2} for large values of n, then Eq. 5.71 says that the wavefunction $\psi(\xi)$ looks like

$$\psi(\xi) = f(\xi)e^{-\frac{\xi^2}{2}} \xrightarrow[\text{large n}]{} e^{\xi^2}e^{-\frac{\xi^2}{2}} = e^{+\frac{\xi^2}{2}}.$$

This positive exponential term increases without limit as $\xi \to \pm\infty$, which means that $\psi(\xi)$ cannot be normalized and is not a physically realizable quantum wavefunction.

But rather than giving up on this approach, you can use this conclusion to take a significant step forward in finding the energy levels of the quantum harmonic oscillator. To take that step, consider how you might prevent $\psi(\xi)$ from blowing up at large positive and negative values of ξ. The answer is to make sure that the series $\sum_n \frac{a_{n+2}}{a_n}$ terminates at some finite value of n, so the series never gets the chance to go like e^{ξ^2} at large values of n.

And what condition can cause this series to terminate? According to Eq. 5.77, the coefficient a_{n+2} equals zero at any value of the energy parameter ϵ for which

$$\frac{2n + (1-\epsilon)}{(n+2)(n+1)} = 0,$$

which means that

$$2n + (1-\epsilon) = 0$$

and

$$\epsilon = 2n + 1. \tag{5.79}$$

This means that the energy parameter ϵ (and therefore the energy E) is quantized, taking on discrete values that depend on the value of n. Denoting this quantization by writing a subscript n, the relationship between E and ϵ (Eq. 5.64) is

$$E_n = \epsilon_n \left(\frac{1}{2}\hbar\omega\right) = (2n + 1)\left(\frac{1}{2}\hbar\omega\right)$$

or

$$E_n = \left(n + \frac{1}{2}\right)\hbar\omega. \tag{5.80}$$

These are the allowed values for the energy of a quantum harmonic oscillator. Just as in the cases of infinite and finite rectangular wells, the quantization of energy and the allowed values of energy come directly from application of the relevant boundary conditions.

You should take a moment to consider these values of the allowed energy. The ground-state ($n = 0$) energy is $E_0 = (1/2)\hbar\omega$, which is exactly what was used as E_{ref} in defining the dimensionless energy parameter ϵ in Eq. 5.62. Note also that the spacing between energy levels of the quantum harmonic oscillator is constant; each energy level E_n is precisely $\hbar\omega$ higher than the adjacent lower level E_{n-1} (as you may recall from the first two sections of this chapter, the spacing between the energy levels of the infinite rectangular well and the finite rectangular well increased with increasing n). So the quantum harmonic oscillator shares some features of the infinite and finite rectangular wells, including quantized energy levels and nonzero ground-state energy, but the variation of the potential with distance from equilibrium results in some significant differences, as well.

With the allowed energies in hand, the next task is to find the corresponding wavefunctions $\psi_n(\xi)$. You can do that using the recursion relation and Eq. 5.71, but you have to think carefully about the limits of the power-series summation.

It's conventional to label the energy levels as E_n, so to distinguish between the index of the energy level and the counter for the terms of the power series, from this point forward the summation index will be labeled as m, making the function $f(\xi)$ look like this:

$$f(\xi) = \sum_{m=0,1,2\ldots} a_m \xi^m. \tag{5.81}$$

Since the recursion equation relates a_{m+2} to a_m, it's helpful to separate this into two series, one with all of the even powers of ξ and the other with all of the odd powers of ξ:

$$f(\xi) = \sum_{m=0,2,4,\ldots} a_m \xi^m + \sum_{m=1,3,5,\ldots} a_m \xi^m. \tag{5.82}$$

You know that the summation terminates (and produces a physically realizable solution) whenever the energy parameter ϵ_n takes on the value $2n+1$. Plugging this into the recursion relation with index m gives

$$a_{m+2} = \frac{2m + (1-\epsilon_n)}{(m+2)(m+1)} a_m = \frac{2m + [1-(2n+1)]}{(m+2)(m+1)} a_m = \frac{2(m-n)}{(m+2)(m+1)} a_m, \tag{5.83}$$

which means the series terminates when $m = n$. So for the first allowed energy level, which has $n = 0$, the energy parameter $\epsilon_0 = 2n + 1 = 1$, and the even series terminates at $m = n = 0$ (meaning that all even terms with $m > n$ are zero). What about the odd series? If you set $a_1 = 0$, the recursion relation will ensure that all higher odd terms will also be zero, which guarantees that the odd series doesn't blow up. So the series for $n = 0$ consists of the single term a_0, and the function $f_0(\xi)$ is

$$f_0(\xi) = \sum_{m=0 \text{ only}} a_m \xi^m = a_0 \xi^0. \tag{5.84}$$

Now consider the first excited ($n = 1$) case. The energy parameter $\epsilon_1 = 2n+1 = 3$ for this first excited state, and the $m-n$ term in Eq. 5.83 causes the odd series to terminate at $m = n = 1$ (so all odd terms with $m > n$ are zero). And to make sure that the even series doesn't blow up, in this case you must set $a_0 = 0$ and the recursion relation will set all higher even terms to zero. So the series for $n = 1$ consists of the single term a_1, and the function $f_1(\xi)$ for the first excited state is

$$f_1(\xi) = \sum_{m=1 \text{ only}} a_m \xi^m = a_1 \xi^1 \tag{5.85}$$

For the second excited state ($n = 2$), the energy parameter $\epsilon_2 = 5$, and the even series terminates at $m = n = 2$. But in this case the counter m can take on the values of 0 and 2, and the recursion relation tells you the ratio

of the coefficients a_2/a_0 and a_4/a_2. For $m = 0$ and $n = 2$, the recursion relation gives

$$a_2 = \frac{2(m-n)}{(m+2)(m+1)}a_m = \frac{2(0-2)}{(0+2)(0+1)}a_0 = -2a_0,$$

and for $m = 2$ and $n = 2$ the recursion relation gives

$$a_4 = \frac{2(m-n)}{(m+2)(m+1)}a_m = \frac{2(2-2)}{(2+2)(2+1)}a_2 = 0,$$

which means the function $f_2(\xi)$ for this second excited state is

$$\begin{aligned} f_2(\xi) &= \sum_{m=0 \text{ and } 2} a_m \xi^m = a_0\xi^0 + a_2\xi^2 \\ &= \left(a_0 + a_2\xi^2\right) = a_0\left(1 - 2\xi^2\right). \end{aligned} \tag{5.86}$$

For the third excited state ($n = 3$), the energy parameter $\epsilon_3 = 7$, and the odd series terminates at $m = n = 3$. In this case the counter m can take on the values of 1 and 3, and the recursion relation tells you the ratio of the coefficients a_3/a_1 and a_5/a_3. For $m = 1$ and $n = 3$, this gives

$$a_3 = \frac{2(m-n)}{(m+2)(m+1)}a_m = \frac{2(1-3)}{(1+2)(1+1)}a_1 = -\frac{2}{3}a_1,$$

and for $m = 3$ and $n = 3$

$$a_5 = \frac{2(m-n)}{(m+2)(m+1)}a_m = \frac{2(3-3)}{(3+2)(3+1)}a_3 = 0.$$

This makes the function $f_3(\xi)$ for this third excited state

$$\begin{aligned} f_3(\xi) &= \sum_{m=1 \text{ and } 3} a_m \xi^m = a_1\xi^1 + a_3\xi^3 \\ &= \left(a_1\xi + a_3\xi^3\right) = a_1\left(\xi - \frac{2}{3}\xi^3\right). \end{aligned} \tag{5.87}$$

For the fourth excited state ($n = 4$), the energy parameter $\epsilon_4 = 9$, and the even series terminates at $m = n = 4$. In this case the counter m can take on the values of 0, 2, and 4 and the recursion relation tells you the ratio of the coefficients a_2/a_0, a_4/a_2, and a_6/a_4. For $m = 0$ and $n = 4$

$$a_2 = \frac{2(m-n)}{(m+2)(m+1)}a_m = \frac{2(0-4)}{(0+2)(0+1)}a_1 = -4a_0,$$

and for $m = 2$ and $n = 4$

$$a_4 = \frac{2(m-n)}{(m+2)(m+1)}a_m = \frac{2(2-4)}{(2+2)(2+1)}a_2$$
$$= \frac{-4}{12}a_2 = -\frac{1}{3}a_2 = \frac{4}{3}a_0.$$

Finally, for $m = 4$ and $n = 4$

$$a_6 = \frac{2(m-n)}{(m+2)(m+1)}a_m = \frac{2(4-4)}{(4+2)(4+1)}a_2 = 0.$$

Hence the function $f_4(\xi)$ for this fourth excited state is

$$f_4(\xi) = \sum_{m=0,2,4} a_m\xi^m = a_0\xi^0 + a_2\xi^2 + a_4\xi^4$$
$$= \left(a_0 + a_2\xi^2 + a_4\xi^4\right) = a_0\left(1 - 4\xi^2 + \frac{4}{3}\xi^4\right). \tag{5.88}$$

For the fifth excited state ($n = 5$), the energy parameter $\epsilon_5 = 11$, and the odd series terminates at $m = n = 5$. In this case the counter m can take on the values of 1, 3, and 5, and the recursion relation tells you the ratio of the coefficients a_3/a_1, a_5/a_3 and a_7/a_5. For $m = 1$ and $n = 5$

$$a_3 = \frac{2(m-n)}{(m+2)(m+1)}a_m = \frac{2(1-5)}{(1+2)(1+1)}a_1 = -\frac{4}{3}a_1,$$

and for $m = 3$ and $n = 5$

$$a_5 = \frac{2(m-n)}{(m+2)(m+1)}a_m = \frac{2(3-5)}{(3+2)(3+1)}a_3$$
$$= \frac{-4}{20}a_3 = -\frac{1}{5}a_3 = \frac{4}{15}a_1.$$

Lastly, for $m = 5$ and $n = 5$

$$a_7 = \frac{2(m-n)}{(m+2)(m+1)}a_m = \frac{2(5-5)}{(5+2)(5+1)}a_5 = 0.$$

Thus the function $f_5(\xi)$ for this fifth excited state is

$$f_5(\xi) = \sum_{m=1,3,5} a_m\xi^m = a_1\xi^1 + a_3\xi^3 + a_5\xi^5$$
$$= \left(a_1\xi + a_3\xi^3 + a_5\xi^5\right) = a_1\left(\xi - \frac{4}{3}\xi^3 + \frac{4}{15}\xi^5\right). \tag{5.89}$$

So these are the first six of the $f_n(\xi)$ functions that produce $\psi_n(\xi)$ when multiplied by the Gaussian exponential factor shown in Eq. 5.71.

And how do these relate to the Hermite polynomials mentioned earlier in the chapter? To see that connection, it helps to collect the f_n functions together and do a little algebra on the argument of each one – specifically, to pull out the constants needed to cause the numerical factor in front of the highest power of ξ for each value of n to be 2^n. That looks like this:

$$
\begin{aligned}
f_0(\xi) &= a_0 = a_0(1) \\
f_1(\xi) &= a_1\xi = \frac{a_1}{2}(2\xi) \\
f_2(\xi) &= a_0\left(1 - 2\xi^2\right) = -\frac{a_0}{2}(4\xi^2 - 2) \\
f_3(\xi) &= a_1\left(\xi - \frac{2}{3}\xi^3\right) = -\frac{a_1}{12}(8\xi^3 - 12\xi) \\
f_4(\xi) &= a_0\left(1 - 4\xi^2 + \frac{4}{3}\xi^4\right) = \frac{a_0}{12}(16\xi^4 - 48\xi^2 + 12) \\
f_5(\xi) &= a_1\left(\xi - \frac{4}{3}\xi^3 + \frac{4}{15}\xi^5\right) = \frac{a_1}{120}(32\xi^5 - 160\xi^3 + 120\xi).
\end{aligned}
$$

The reason for this manipulation is to make it easy to compare the $f_n(\xi)$ functions with the Hermite polynomials. If you look up those polynomials in a physics text or online, you're likely[8] to find these expressions:

$$
\begin{aligned}
H_0(\xi) &= 1 & H_1(\xi) &= 2\xi \\
H_2(\xi) &= 4\xi^2 - 2 & H_3(\xi) &= 8\xi^3 - 12\xi \\
H_4(\xi) &= 16\xi^4 - 48\xi^2 + 12 & H_5(\xi) &= 32\xi^5 - 160\xi^3 + 120\xi.
\end{aligned}
$$

Comparing the $f_n(\xi)$ functions to the Hermite polynomials $H_n(\xi)$, you can see that they're identical except for the constant factors involving a_0 or a_1 in $f_n(\xi)$. Calling those constants A_n, Eq. 5.71 gives the wavefunction $\psi_n(\xi)$ as

$$\psi_n(\xi) = f_n e^{-\frac{\xi^2}{2}} = A_n H_n(\xi) e^{-\frac{\xi^2}{2}}, \tag{5.90}$$

and the constants will be determined by normalizing the wavefunctions $\psi_n(\xi)$, which is the next task.

Before getting to that, take a look at the terms in Eq. 5.90. As promised, the quantum harmonic oscillator wavefunctions are comprised of the product

[8] If you come across a list of Hermite polynomials with different numerical factors (such as unity rather than 2^n in front of the highest power of ξ), you may be looking at the "probabalist's" version rather than the "physicist's" version of Hermite polynomials, which differ only in the scaling factor.

of Hermite polynomials (H_n) and a Gaussian exponential ($e^{-\xi^2/2}$). It's the Gaussian term that causes the wavefunction $\psi(\xi)$ to decrease toward zero as ξ goes to $\pm\infty$, providing the spatial localization needed for normalization.

To accomplish that normalization, set the integrated probability density over all space to unity. For $\psi_n(x)$, the integration is over x:

$$\int_{-\infty}^{\infty} \psi^*(x)\psi(x)dx = 1 \tag{5.91}$$

and Eq. 5.66 relates dx to $d\xi$, so

$$\int_{-\infty}^{\infty} \psi^*(x)\psi(x)dx = \sqrt{\frac{\hbar}{m\omega}} \int_{-\infty}^{\infty} \psi^*(\xi)\psi(\xi)d\xi = 1, \tag{5.92}$$

which means

$$\sqrt{\frac{\hbar}{m\omega}} \int_{-\infty}^{\infty} \left[A_n H_n(\xi) e^{-\frac{\xi^2}{2}}\right]^* \left[A_n H_n(\xi) e^{-\frac{\xi^2}{2}}\right] d\xi = 1 \tag{5.93}$$

or

$$\sqrt{\frac{\hbar}{m\omega}} |A_n|^2 \int_{-\infty}^{\infty} |H_n(\xi)|^2 e^{-\xi^2} d\xi = 1. \tag{5.94}$$

This integral looks nasty, but mathematicians working on Weber equations and Hermite polynomials have given us a very handy integral identity:

$$\int_{-\infty}^{\infty} |H_n(\xi)|^2 e^{-\xi^2} d\xi = 2^n n! \left(\pi^{\frac{1}{2}}\right), \tag{5.95}$$

which is exactly what we need. Inserting this expression into Eq. 5.94 yields

$$\sqrt{\frac{\hbar}{m\omega}} |A_n|^2 \left[2^n n! \left(\pi^{\frac{1}{2}}\right)\right] = 1$$

$$|A_n|^2 = \sqrt{\frac{m\omega}{\hbar}} \frac{1}{2^n n! \left(\pi^{\frac{1}{2}}\right)},$$

and taking the square root gives the normalization constant A_n:

$$|A_n| = \left(\frac{m\omega}{\hbar}\right)^{\frac{1}{4}} \frac{1}{\sqrt{2^n n! \left(\pi^{\frac{1}{2}}\right)}} = \left(\frac{m\omega}{\pi\hbar}\right)^{\frac{1}{4}} \frac{1}{\sqrt{2^n n!}}. \tag{5.96}$$

With A_n in hand, you can write the wavefunction $\psi_n(\xi)$ as

$$\psi_n(\xi) = \left(\frac{m\omega}{\pi\hbar}\right)^{\frac{1}{4}} \left(\frac{1}{\sqrt{2^n n!}}\right) H_n(\xi) e^{-\frac{\xi^2}{2}}. \tag{5.97}$$

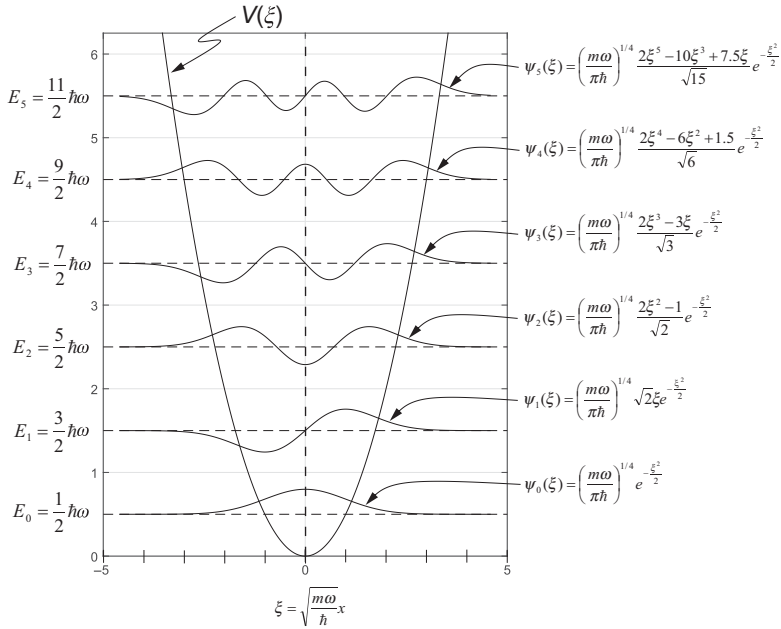


Figure 5.26 Quantum harmonic oscillator wavefunctions $\psi_n(\xi)$.

These wavefunctions $\psi_n(\xi)$ for the six lowest energy levels of the quantum harmonic oscillator are shown in Fig. 5.26. As in the case of the rectangular wells, the lowest-energy (ground-state) wavefunction is even with respect to the center of the well ($x = 0$), and higher-energy wavefunctions alternate between odd and even parity. Like the solutions for the finite rectangular well, the harmonic-oscillator wavefunctions are oscillatory in the classically allowed region and exponentially decaying in the classically forbidden regions. In the classically allowed region, the curvature of the wavefunction increases with increasing energy, so higher-energy wavefunctions have more cycles between the classical turning points – specifically, ψ_n has one more (partial) half-cycle and one more node than ψ_{n-1}.

The probability densities $P_{den}(\xi) = \psi_n^*(\xi)\psi_n(\xi)$ for the six lowest-energy wavefunctions of the harmonic oscillator are shown in Fig. 5.27. These plots make it clear that at low energies (small values of n), the behavior of the quantum harmonic oscillator differs significantly from the classical case. For example, for a particle in the ground state, the energy is $\hbar\omega/2$, and a position measurement is most likely to yield a value near $x = 0$. Additionally,

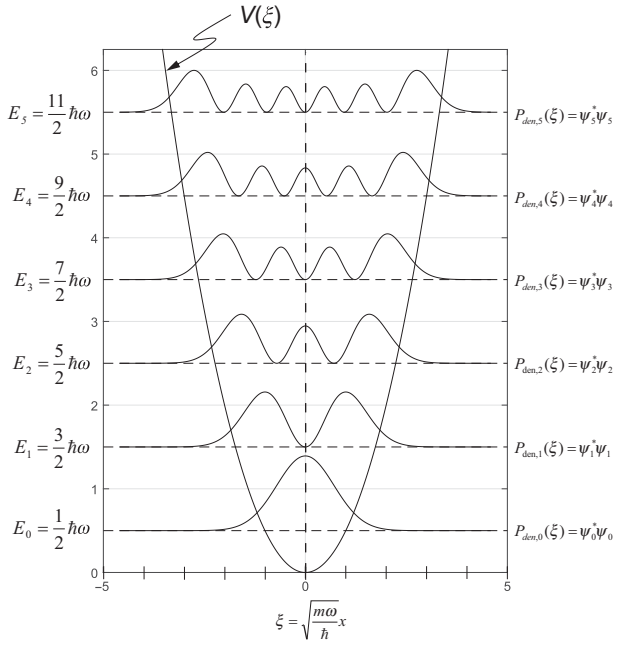


Figure 5.27 Quantum harmonic oscillator probability densities.

each of the excited states $\psi_n(\xi)$ has n locations with zero probability within the classically allowed region. However, if you look closely you can see that as n increases, the probability of a position measurement producing a result near the classical turning points increases, so the behavior of the quantum harmonic oscillator does begin to resemble that of the classical case at large values of n, as required by the Correspondence Principle described in Section 4.1.

You should also bear in mind that these wavefunctions are the eigenfunctions of the Hamiltonian operator achieved by separation of variables, so they represent stationary states for which the expectation values of observables such as position, momentum, and energy do not change over time. To determine the behavior of particles in other states (all of which can be synthesized as weighted combinations of these eigenstates), you must include the time function $T(t)$, which makes $\Psi_n(x,t)$

$$\Psi_n(x,t) = \left(\frac{m\omega}{\pi\hbar}\right)^{\frac{1}{4}} \left(\frac{1}{\sqrt{2^n n!}}\right) H_n\left(\sqrt{\frac{m\omega}{\hbar}}x\right) e^{-\frac{m\omega}{2\hbar}x^2} e^{-i\left(n+\frac{1}{2}\right)\omega t}. \qquad (5.98)$$

Knowing the allowed energy levels E_n and wavefunctions $\Psi_n(x,t)$ allows you to determine the behavior of the quantum harmonic oscillator over space and time. That behavior includes the expectation values of observables such as position x and momentum p, as well as the square magnitudes of those quantities and the resulting uncertainties (you can see examples of that in the chapter-end problems and online solutions).

Thus the analytic approach has provided the tools you need to analyze this important configuration. But you may also find it useful to understand the algebraic approach to finding the energy levels and wavefunctions for the quantum harmonic oscillator, so that's the subject of the remainder of this chapter.

The algebraic approach involves a dimensionless version of the time-independent Schrödinger equation, written using dimensionless versions of the position and momentum operators $\widehat{X}$ and $\widehat{P}$. To see how that works, start by defining a momentum reference value p_{ref} using

$$\frac{p_{ref}^2}{2m} = E_{ref} = \frac{1}{2}\hbar\omega$$
$$p_{ref} = \sqrt{\frac{2m\hbar\omega}{2}} = \sqrt{m\hbar\omega},$$

and use this expression to produce a dimensionless version of momentum called $\mathcal{P}$ as

$$\mathcal{P} = \frac{p}{p_{ref}} = \frac{p}{\sqrt{m\hbar\omega}}, \tag{5.99}$$

or, writing momentum p in terms of $\mathcal{P}$,

$$p = \mathcal{P}(p_{ref}). \tag{5.100}$$

To produce a dimensionless version of the TISE, write energy E in terms of dimensionless energy ϵ, position x in terms of dimensionless position ξ, and momentum p in terms of dimensionless momentum $\mathcal{P}$. Start with the TISE from Chapter 3:

$$-\frac{\hbar^2}{2m}\frac{d^2[\psi(x)]}{dx^2} + V[\psi(x)] = E[\psi(x)], \tag{3.40}$$

which can be written in terms of the momentum operator $\widehat{P}$ and position operator $\widehat{X}$ for the quantum harmonic oscillator as

$$\left[\frac{\widehat{P}^2}{2m} + \frac{1}{2}m\omega^2\widehat{X}^2\right][\psi(x)] = E[\psi(x)].$$

Using dimensionless operators $\hat{\mathcal{P}} = \widehat{P}/p_{ref}$ and $\hat{\xi} = \widehat{X}/x_{ref}$ makes this

$$\left[\frac{[\hat{\mathcal{P}}(p_{ref})]^2}{2m} + \frac{1}{2}m\omega^2[\hat{\xi}(x_{ref})]^2\right][\psi(\xi)] = \epsilon(E_{ref})[\psi(\xi)]$$

or

$$\left[\frac{(\hat{\mathcal{P}}\sqrt{m\hbar\omega})^2}{2m} + \frac{1}{2}m\omega^2\left(\hat{\xi}\sqrt{\frac{\hbar}{m\omega}}\right)^2\right][\psi(\xi)] = \epsilon\left(\frac{1}{2}\hbar\omega\right)[\psi(\xi)]$$

$$\left[\hat{\mathcal{P}}^2\frac{\hbar\omega}{2} + \hat{\xi}^2\frac{\hbar\omega}{2}\right][\psi(\xi)] = \epsilon\left(\frac{\hbar\omega}{2}\right)[\psi(\xi)].$$

Removing the common factor of $\hbar\omega/2$ gives a straightforward version of the TISE:

$$\left[\hat{\mathcal{P}}^2 + \hat{\xi}^2\right][\psi(\xi)] = \epsilon[\psi(\xi)]. \tag{5.101}$$

The algebraic approach to solving this equation begins with the definition of two new operators, which are combinations of the dimensionless position and momentum operators. The first of these new operators is

$$\hat{a}^\dagger = \frac{1}{\sqrt{2}}(\hat{\xi} - i\hat{\mathcal{P}}) \tag{5.102}$$

and the second is

$$\hat{a} = \frac{1}{\sqrt{2}}(\hat{\xi} + i\hat{\mathcal{P}}). \tag{5.103}$$

In some texts, you'll see these operators written as $\hat{a}^+$ and $\hat{a}^-$. The reason for that notation and for using this combination of operators with a leading factor of $1/\sqrt{2}$ will become clear when you see how these operators act on the wavefunctions of the quantum harmonic oscillator.

Each of these two operators will prove useful once any wavefunction solution $\psi_n(\xi)$ is known, but it's their product that can help you find those wavefunctions. That product is

$$\begin{aligned}\hat{a}^\dagger\hat{a} &= \frac{1}{\sqrt{2}}(\hat{\xi} - i\hat{\mathcal{P}})\frac{1}{\sqrt{2}}(\hat{\xi} + i\hat{\mathcal{P}}) \\ &= \frac{1}{2}\left(\hat{\xi}^2 + i\hat{\xi}\hat{\mathcal{P}} - i\hat{\mathcal{P}}\hat{\xi} + \hat{\mathcal{P}}^2\right).\end{aligned}$$

As you can see, the terms $\hat{\mathcal{P}}^2 + \hat{\xi}^2$ on the left side of the TISE (Eq. 5.101) are present in this expression, along with two cross terms that involve both

$\hat{\xi}$ and $\hat{\mathcal{P}}$. Now look at what happens if you factor out the imaginary unit i from those cross terms:

$$i\hat{\xi}\hat{\mathcal{P}} - i\hat{\mathcal{P}}\hat{\xi} = i(\hat{\xi}\hat{\mathcal{P}} - \hat{\mathcal{P}}\hat{\xi}) = i[\hat{\xi}, \hat{\mathcal{P}}], \tag{5.104}$$

in which $[\hat{\xi}, \hat{\mathcal{P}}]$ represents the commutator of the operators $\hat{\xi}$ and $\hat{\mathcal{P}}$. This makes the product $\hat{a}^\dagger\hat{a}$ look like this:

$$\hat{a}^\dagger\hat{a} = \frac{1}{2}\left(\hat{\xi}^2 + \hat{\mathcal{P}}^2 + i[\hat{\xi}, \hat{\mathcal{P}}]\right). \tag{5.105}$$

This can be simplified by writing the commutator in terms of $\widehat{X}$ and $\widehat{P}$:

$$i[\hat{\xi}, \hat{\mathcal{P}}] = i\left[\frac{\widehat{X}}{x_{ref}}, \frac{\widehat{P}}{p_{ref}}\right] = \frac{i}{x_{ref}p_{ref}}[\widehat{X}, \widehat{P}]$$

or

$$i[\hat{\xi}, \hat{\mathcal{P}}] = \frac{i}{\sqrt{\frac{\hbar}{m\omega}}\sqrt{m\hbar\omega}}[\widehat{X}, \widehat{P}] = \frac{i}{\hbar}[\widehat{X}, \widehat{P}].$$

Recall from Chapter 4 that the canonical commutation relation (Eq. 4.68) tells you that $[\widehat{X}, \widehat{P}] = i\hbar$, which means

$$i[\hat{\xi}, \hat{\mathcal{P}}] = \frac{i}{\hbar}[i\hbar] = -1. \tag{5.106}$$

Plugging this into Eq. 5.105 gives

$$\hat{a}^\dagger\hat{a} = \frac{1}{2}\left(\hat{\xi}^2 + \hat{\mathcal{P}}^2 - 1\right)$$

or

$$\left(\hat{\xi}^2 + \hat{\mathcal{P}}^2\right) = 2\hat{a}^\dagger\hat{a} + 1. \tag{5.107}$$

This makes the TISE (Eq. 5.101)

$$\left[\mathcal{P}^2 + \xi^2\right][\psi(\xi)] = (2\hat{a}^\dagger\hat{a} + 1)[\psi(\xi)] = \epsilon[\psi(\xi)]$$

or

$$2\hat{a}^\dagger\hat{a}[\psi(\xi)] = (\epsilon - 1)[\psi(\xi)].$$

Plugging the definitions of $\hat{a}^\dagger$ and $\hat{a}$ into this equation gives

$$2\left[\frac{1}{\sqrt{2}}(\hat{\xi} - i\hat{\mathcal{P}})\frac{1}{\sqrt{2}}(\hat{\xi} + i\hat{\mathcal{P}})\right][\psi(\xi)] = (\epsilon - 1)[\psi(\xi)]$$

or

$$(\hat{\xi} - i\hat{\mathcal{P}})(\hat{\xi} + i\hat{\mathcal{P}})[\psi(\xi)] = (\epsilon - 1)[\psi(\xi)]. \tag{5.108}$$

One way that this equation can be satisfied is for the dimensionless energy parameter ϵ to equal unity while $(\hat{\xi} + i\hat{\mathcal{P}})\psi(\xi)$ equals zero.

If $\epsilon = 1$, the total energy is

$$E = \epsilon E_{ref} = (1)\left(\frac{\hbar\omega}{2}\right) = \frac{1}{2}\hbar\omega,$$

in agreement with the ground-state energy level E_0 determined by the power-series approach.

The wavefunction $\psi_0(\xi)$ corresponding to this energy level can be found by setting the term $(\hat{\xi} + i\hat{\mathcal{P}})\psi(\xi)$ on the other side of Eq. 5.108 to zero. To see how that works, use Eq. 5.66 to write the momentum operators $\widehat{P}$ and $\hat{\mathcal{P}}$ as

$$\widehat{P} = -i\hbar\frac{d}{dx} = -i\hbar\frac{d}{\sqrt{\frac{\hbar}{m\omega}}d\xi} = -i\sqrt{m\hbar\omega}\frac{d}{d\xi}$$

and

$$\hat{\mathcal{P}} = \frac{\widehat{P}}{p_{ref}} = \frac{-i\sqrt{m\hbar\omega}}{\sqrt{m\hbar\omega}}\frac{d}{d\xi} = -i\frac{d}{d\xi}.$$

This means that if $(\hat{\xi} + i\hat{\mathcal{P}})\psi(\xi) = 0$, then

$$\begin{aligned}(\hat{\xi} + i\hat{\mathcal{P}})\psi(\xi) = \left[\xi + i\left(-i\frac{d}{d\xi}\right)\right]\psi(\xi) &= 0\\ \left[\xi + \frac{d}{d\xi}\right]\psi(\xi) &= 0\\ \frac{d\psi(\xi)}{d\xi} &= -\xi\psi(\xi).\end{aligned} \tag{5.109}$$

The solution to this equation is $\psi(\xi) = Ae^{-\frac{\xi^2}{2}}$, and normalizing gives $A = (\frac{m\omega}{\pi\hbar})^{1/4}$ (if you need help getting that result, see the chapter-end problems and online solutions).

Hence the algebraic approach gives the lowest-energy eigenfunction

$$\psi(\xi) = \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} e^{-\frac{\xi^2}{2}},$$

exactly as found for $\psi_0(\xi)$ using the analytic approach.

So the operator product $\hat{a}^\dagger\hat{a}$ has proven useful in finding the lowest-energy solution to the Schrödinger equation for the quantum harmonic oscillator. But

as mentioned, the operators $\hat{a}^\dagger$ and $\hat{a}$ are also useful individually. You can see this by applying the $\hat{a}^\dagger$ operator to the ground-state wavefunction:

$$\begin{aligned}
\hat{a}^\dagger\psi_0(\xi) &= \frac{1}{\sqrt{2}}(\hat{\xi} - i\hat{\mathcal{P}})\left[\left(\frac{m\omega}{\pi\hbar}\right)^{1/4} e^{-\frac{\xi^2}{2}}\right] \\
&= \frac{\xi}{\sqrt{2}}\left[\left(\frac{m\omega}{\pi\hbar}\right)^{1/4} e^{-\frac{\xi^2}{2}}\right] + \frac{-i}{\sqrt{2}}\left\{-i\frac{d}{d\xi}\left[\left(\frac{m\omega}{\pi\hbar}\right)^{1/4} e^{-\frac{\xi^2}{2}}\right]\right\} \\
&= \frac{1}{\sqrt{2}}\left(\frac{m\omega}{\pi\hbar}\right)^{1/4}\left[\xi e^{-\frac{\xi^2}{2}} - \frac{d}{d\xi}\left(e^{-\frac{\xi^2}{2}}\right)\right] \\
&= \frac{1}{\sqrt{2}}\left(\frac{m\omega}{\pi\hbar}\right)^{1/4}\left[\xi e^{-\frac{\xi^2}{2}} - \frac{-2\xi}{2}e^{-\frac{\xi^2}{2}}\right] \\
&= \frac{1}{\sqrt{2}}\left(\frac{m\omega}{\pi\hbar}\right)^{1/4}\left[\xi e^{-\frac{\xi^2}{2}} + \xi e^{-\frac{\xi^2}{2}}\right] \\
&= \frac{1}{\sqrt{2}}\left(\frac{m\omega}{\pi\hbar}\right)^{1/4}\left[2\xi e^{-\frac{\xi^2}{2}}\right] = \left(\frac{m\omega}{\pi\hbar}\right)^{1/4}\left[\sqrt{2}\xi e^{-\frac{\xi^2}{2}}\right] \\
&= \psi_1(\xi).
\end{aligned}$$

So applying the $\hat{a}^\dagger$ operator to the ground-state wavefunction $\psi_0(\xi)$ produces the wavefunction $\psi_1(\xi)$ of the first excited state. For this reason, $\hat{a}^\dagger$ is known as a "raising" operator – each time it's applied to a wavefunction $\psi_n(\xi)$ of the quantum harmonic oscillator, it produces a wavefunction proportional to the wavefunction with the next higher quantum number $\psi_{n+1}(\xi)$. For the raising operator, the constant of proportionality is $\sqrt{n+1}$, so

$$\hat{a}^\dagger\psi_n(\xi) = \sqrt{n+1}\psi_{n+1}(\xi). \tag{5.110}$$

When the raising operator is applied to the ground state, this means $\hat{a}^\dagger\psi_0(\xi) = \sqrt{0+1}\psi_{0+1}(\xi) = \psi_1(\xi)$.

As you may have surmised, the operator $\hat{a}$ performs the complementary function, producing a wavefunction proportional to the wavefunction with the quantum number lowered by one. Hence $\hat{a}$ is called a "lowering operator," and for the lowering operator, the constant of proportionality is $\sqrt{n}$. Thus

$$\hat{a}\psi_n(\xi) = \sqrt{n}\psi_{n-1}(\xi). \tag{5.111}$$

This is why $\hat{a}^\dagger$ and $\hat{a}$ are known as ladder operators; they allow you to "climb" up or down the wavefunctions of the quantum harmonic oscillator. These wavefunctions have different energy levels, so some texts refer to the ladder

operators as "creation" and "annihilation" operators – each step up creates and each set down destroys one quantum ($\frac{1}{2}\hbar\omega$) of energy.

If you'd like to get some experience using a ladder operator and applying the other mathematical concepts and techniques described in this chapter, take a look at the problems in the final section. As always, you can find full interactive solutions to each of these problems on the book's website.

5.4 Problems

1. Show that a global phase factor such as $e^{i\theta}$ that applies equally to all component wavefunctions ψ_n that are superposed to produce wavefunction $\psi(x)$ cannot affect the probability density, but that the relative phase of the component wavefunctions *does* have an effect on the probability density.

2. For a particle in the ground state of an infinite rectangular potential well, use the position operator $\widehat{X}$ and the momentum operator $\widehat{P}$ to find the expectation values $\langle x\rangle$ and $\langle p\rangle$. Then use the square of the position and momentum operators to find $\langle x^2\rangle$ and $\langle p^2\rangle$.

3. Use your results from the previous problem to find the uncertainties Δx and Δp and show that the Heisenberg Uncertainty principle is satisfied.

4. If a particle in an infinite rectangular potential well has wavefunction $\psi(x) = \frac{1}{2}\psi_1(x) + \frac{3i}{4}\psi_2(x) + \frac{\sqrt{3}}{4}\psi_3(x)$, in which the functions ψ_n are given by Eq. 5.9,

a) What are the possible results of a measurement of the particle's energy, and what is the probability of each result?

b) Find the expectation value of the energy for this particle.

5. Determine the probability of finding a particle in the region between $x = 0.25a$ and $x = 0.75a$ in an infinite rectangular potential well of width a centered on $x = a/2$ if the particle is in the first excited state and if the particle is in the second excited state.

6. Derive the expression for $\tilde{\phi}(p)$ given by Eq. 5.16, and use that result to derive the expression for $P_{den}(p)$ given by Eq. 5.17.

7. Find the expectation values $\langle x\rangle$, $\langle p\rangle$, $\langle x^2\rangle$, and $\langle p^2\rangle$ for a particle in the ground state of a quantum harmonic oscillator.

8. Use your results from the previous problem to find the uncertainties Δx and Δp and show that the Heisenberg Uncertainty principle is satisfied.

9. Show that the normalization constant $A = \left(\frac{m\omega}{\pi\hbar}\right)^{1/4}$ is correct for the solution to Eq. 5.109 for the ground state of the quantum harmonic oscillator.

10. a) Apply the lowering operator $\hat{a}$ to $\psi_2(x)$ for the quantum harmonic oscillator and use the result to find $\psi_1(x)$.

b) Show that the position operator $\widehat{X}$ and the momentum operator $\widehat{P}$ can be written in terms of the ladder operators $\hat{a}^\dagger$ and $\hat{a}$ as

$$\widehat{X} = \sqrt{\frac{\hbar}{2m\omega}}(\hat{a}^\dagger + \hat{a})$$

and

$$\widehat{P} = i\sqrt{\frac{\hbar m\omega}{2}}(\hat{a}^\dagger - \hat{a}).$$

References

[1] Cohen-Tannoudji, C., B. Diu, and F. Laloë, *Quantum Mechanics*, John Wiley & Sons, 1977.

[2] Goswami, A., *Quantum Physics*, William C. Brown, 1990.

[3] Griffiths, D., *Introduction to Quantum Mechanics*, Pearson Prentice-Hall, 2005.

[4] Marshman, E., and C. Singh, "Review of student difficulties in upper-level quantum mechanics," *Phys. Rev. ST Phys. Educ. Res.*, 11(2), 2015, 020117.

[5] McMahon, D., *Quantum Mechanics Demystified*, McGraw-Hill, 2014.

[6] Messiah, A., *Quantum Mechanics*, Dover, 2014.

[7] Morrison, M., *Understanding Quantum Physics*, Prentice-Hall, 1990.

[8] Phillips, A. C., *Introduction to Quantum Mechanics*, John Wiley & Sons, 2003.

[9] Rapp, D., *Quantum Mechanics*, CreateSpace Independent Publishing Platform, 2013.

[10] Susskind, L., and A. Friedman, *Quantum Mechanics: The Theoretical Minimum*, Basic Books, 2014.

[11] Townsend, J., *A Modern Approach to Quantum Mechanics*, University Science Books, 2012.

[12] Zettili, N., *Quantum Mechanics: Concepts and Applications*, John Wiley & Sons, 2009.

Index

abstract space, 14
abstract vectors, 14
adjoint, 44
algebraic approach to harmonic oscillator, 212
allowed energies
  infinite rectangular well, 152
allowed region, 73
analytical addition of vectors, 3
angular frequency, 64

basis system
  natural, 9
  orthonormal, 39
  standard, 9, 54
basis vectors, 2
Bohr, Niels, 56
Born rule, 60, 77, 96
Born rule in Copenhagen interpretation, 97
Born, Max, 60, 77, 96
boundary between regions, 111
bra, 10
  as dual to ket, 12
  mapping vectors to scalars, 12
  as row vector, 11

canonical commutation relation, 143
Cartesian coordinate system, 2
classically forbidden, 105
closure relation, 52
commutation, 41
complementarity in Copenhagen interpretation, 97
complete set, 7
completeness, 52
complex conjugate, 20
complex numbers, 18
complex plane, 19, 117
components of a vector, 2
conjugate variables, 124
conjugate-transpose, 44
coordinate system
  Cartesian, 2
  spherical, 9
Copenhagen interpretation, 96
Correspondence Principle
  for harmonic oscillator, 210
correspondence principle in Copenhagen interpretation, 97
covector, 12
creation and annihilation operators, 216
curvature
  proportional to wavefunction, 103
  related to $E - V$, 107
  in Schrödinger equation, 73

de Broglie relation, 65, 133, 152
de Broglie, Louis, 64
decay constant, 109, 171
degenerate eigenfunctions, 48
del symbol, 89
del-squared operator, 85
depth of finite rectangular well, 169
derivative operator, 36
diffusion equation, 74
  three-dimensional, 86
Dirac delta function, 103
  definition, 130

Dirac notation, 8
  operators in, 37
Dirac, Paul, 10
directional indicators, 2
Dirichlet conditions, 112
dispersion relation, 156
distribution, 103
divergence, 88
domains, 124
dot product
  generalized, 10
  in planewave function, 83
  of vectors, 5
dual space, 12

eigenfunction, 36
  of the momentum operator, 140
  of the position operator, 140
eigenket, 37
eigenvalue, 35
eigenvalue equation, 80
eigenvector, 35
Einstein, Albert, 64
energy
  photon, 64
ensemble of systems, 56
escape to infinity by free particle, 111
Euler relation, 108, 112
  inverse, 121, 151
evanescent region, 169
even function, 154
excited states of infinite rectangular well, 154
expanding a vector, 2
expectation values, 56

Feynman, Richard, 64
finite potential well
  weak, 186
finite rectangular potential well, 168
finite rectangular well
  depth, 169
force related to potential energy, 147
Fourier, Jean-Baptiste Joseph, 29
Fourier synthesis, 124
Fourier theory, 29, 111
Fourier transform, 101, 111
  between $\psi(x)$ and $\phi(k)$, 69
  in Dirac notation, 123
  pairs, 124
free particle, 99, 111
frequency, 64
functions as abstract vectors, 14

Gaussian function characteristics, 134
Gaussian wave packet, 134
Gaussian wavefunction
  two-dimensional, 93
generalized, 10
  coordinate, 14
  function, 103
global phase factor, 153
ground state
  infinite rectangular well, 153
gradient, 88
Gram–Schmidt procedure, 48
graphical addition of vectors, 3
guiding wave, 96

Hamiltonian operator, 69
  three-dimensional, 92
harmonic functions, 25
harmonic oscillator, 193
  algebraic approach, 196
  analytic approach, 196
  harmonic approach, 212
Heisenberg Uncertainty principle, 133, 138
Heisenberg, Werner, 95
Hermite polynomials, 198, 207
  probabilist's vs. physicist's, 207
Hermitian conjugate, 44
Hermitian operator, 43
  real eigenvalues, 48
  self-adjoint, 46
Hilbert space, 15
Hilbert, David, 15
Hooke's Law, 194

identity operator, 51, 123
imaginary numbers, 19
imaginary unit, 19, 72
  as rotation operator, 72
infinite rectangular potential well, 147
  allowed energies, 152
  ground state, 153
information content in Copenhagen
    interpretation, 97
inner product, 10
  of functions, 17
  space, 15
  of vectors, 5

inverse Euler relation, 121, 151
inverse Fourier tranform, 112

ket, 10
  definition, 11
  independent of basis, 12
kinetic energy
  classical equation, 65
  negative, 105
Kronecker delta, 28

L'Hôpital's rule, 126, 144
ladder operator, 196
Laplacian operator, 85
  spherical coordinates, 92
linear functional, 12
linear operator, 33
linear space, 11
linear transformations, 33
lowering operator, 215

magnitude
  of a complex number, 19
  of a vector, 2
main lobe of sinc function, 126
mass term in Schrödinger equation, 73
Matrix mechanics, 95
matter wave, 64, 96
Maxwell, James Clerk, 64
measurement results in Copenhagen interpretation, 97
momentum of electromagnetic wave, 64
momentum operator, 70
momentum operator eigenfunctions, 140
momentum-energy relation, 65
monochromatic plane wave, 128
multiplying a vector by a scalar, 3

nabla symbol, 89
natural basis, 9
negative kinetic energy, 105
Newton's Second Law, 77
Niels Bohr Institute, 96
norm of a vector, 2
normal modes, 156
normalizable functions, 18
normalization, 28
number lines, 19

odd function, 154
one-form, 12
operator
  3-D Hamiltonian, 92
  adjoint of, 44
  creation and annihilation, 216
  derivative, 36
  in Dirac notation, 37
  Hamiltonian, 69
  Hermitian, 43
  identity, 51, 123
  ladder, 196
  Laplacian, 85
  linear, 33
  lowering, 215
  momentum, 70
  projection, 49, 123
  quantum, 32
  raising, 215
  sandwiching, 38
  second-derivative, 36
  total-energy, 69
orthogonal functions, 22
orthogonality of harmonic functions, 25
orthogonality relations, 113
orthonormal basis system, 39
orthonormal basis vectors, 28
orthonormal vectors, 6

particle trapped in potential well, 147
path integral, 71
phase factor
  global, 153
phasors, 116
photon energy, 64
piecewise constant potential, 110
Planck constant, 64
  dimensions of, 72
  modified, 72
Planck, Max, 64
Planck–Einstein relation, 65
plane wave, 66, 82
  monochromatic, 124, 128
position operator eigenfunctions, 140
potential energy
  quadratic, 194
  reference level, 147, 168
  term in Schrödinger equation, 73
potential vs. potential energy, 73

potential well
  finite rectangular, 168
  infinite rectangular, 147
probability
  amplitude, 99
  density function, 77
  distribution width, 137
  flow, 71
projection operator, 49, 123
Pythagorean theorem, 19

quadratic potential energy, 194
quantized wavenumber, 151
quantum operator, 32
quantum state, 13
quantum state collapse, 101
quantum state vs. wavefunction, 98
quantum wavefunctions characteristics, 102

radiation pressure, 64
raising operator, 215
real numbers, 19
rectangular potential well
  finite, 168
  infinite, 147
reference level for potential energy, 147, 168

sandwiching an operator, 38
scalar product of vectors, 4
Schrödinger equation, 63
  expanded view, 71
  meaning of, 71
  origin of, 64
  three-dimensional, 81
  time-dependent, 67
  time-independent, 78, 80
Schrödinger, Erwin, 95
second-derivative operator, 36
separation of variables, 78
sidelobes of sinc function, 126
sinc function, 126
single-valued function, 102
smooth function, 102
spatial frequency, 112
spectral decomposition, 29, 111
spectrum
  of eigenvalues, 99
  of a function, 29
speed of light, 64
spherical coordinate system, 9
  Laplacian operator in, 92
spring constant, 194
square integrable, 18, 103
square well, 147
square-pulse function, 31
standard basis, 9
standard basis system, 54
standard deviation, 137
stationary states, 81, 163
superposition, 68
surfaces of constant phase, 83

three-dimensional Schrödinger equation, 81
time evolution in Copenhagen interpretation, 97
time-independent Schrödinger equation, 78, 80
total energy, 65
total energy operator, 69
  three-dimensional, 92
transcendental equation, 173
transpose conjugate, 44
transpose of a matrix, 44
turning points, 194

unallowed region, 73
uncertainty principle, 124
  in Copenhagen interpretation, 97
uncertainty related to expectation values, 59
uncertainty relation for Gaussian wavefunctions, 137
unit vectors, 2
units of Ψ, 72

variance, 59
vector, 2
  adding, 3
  bra, 10
  complete set, 7
  complex, 18
  components, 2
  dot product, 5
  expanding in a basis, 2
  independence from basis systems, 2
  inner product, 5, 10
  ket, 10
  magnitude of, 2
  multiplying by a scalar, 3

natural basis system, 9
norm of, 2
orthonormal, 6
scalar product, 4
single-column matrix, 8
standard basis system, 9
unit, 2
vector space, 11

Wave mechanics, 95
wave packet, 111, 123
wave-particle duality, 65
wavefunction
position and momentum, 132
units of, 72
wavefunction collapse, 152
in Copenhagen interpretation, 97
wavefunction vs. quantum state, 98
wavefunctions, 13
wavenumber
definition of, 65
quantized, 151
spectrum, 112
three-dimensional, 82
weak finite potential well, 186
Weber equations, 198
well-behaved function, 102
in Fourier theory, 112

zero-crossing of sinc function, 126
zero-point energy, 152